SECOND EDITION

Theory of Knowledge

KEITH LEHRER

Dimensions of Philosophy Series

THEORY OF KNOWLEDGE

This page intentionally left blank

DIMENSIONS OF PHILOSOPHY SERIES
Norman Daniels and Keith Lehrer, Editors

Theory of Knowledge, Second Edition, Keith Lehrer

Philosophy of Biology, Second Edition, Elliott Sober

Contemporary Continental Philosophy, Robert D'Amico

Philosophical Ethics, Stephen Darwall

Normative Ethics, Shelly Kagan

Introduction to Marx and Engels: A Critical Reconstruction, Second Edition, Richard Schmitt

Political Philosophy, Jean Hampton

Philosophy of Mind, Jaegwon Kim

Philosophy of Social Science, Second Edition, Alexander Rosenberg

Philosophy of Education, Nel Noddings

Philosophy of Law: An Introduction to Jurisprudence, Revised Edition, Jeffrie G. Murphy and Jules L. Coleman

Philosophy of Physics, Second Edition, forthcoming, Lawrence Sklar

Metaphysics, Second Edition, forthcoming, Peter van Inwagen

Philosophy of Science, forthcoming, Lawrence Sklar

Philosophy of Cognitive Science, forthcoming, Barbara Von Eckardt

This page intentionally left blank

THEORY OF KNOWLEDGE

Second Edition

Keith Lehrer
University of Arizona

A Member of the Perseus Books Group

Dimensions of Philosophy Series

All rights reserved. Printed in the United States of America. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Copyright © 2000 by Westview Press, A Member of the Perseus Books Group

Published in 2000 in the United States of America by Westview Press, 5500 Central Avenue, Boulder, Colorado 80301-2877, and in the United Kingdom by Westview Press, 12 Hid's Copse Road, Cumnor Hill, Oxford OX2 9JJ

Visit us on the World Wide Web at www.westviewpress.com

Library of Congress Cataloging-in-Publication Data
Lehrer, Keith.
Theory of knowledge / Keith Lehrer.—2nd ed.
p. cm.—(Dimensions of philosophy series)
Includies bibliographical references and index.
ISBN 0-8133-9053-2
1. Knowledge, Theory of. I. Title. II. Series.

BD161 .L368 2000
121—dc21 00-043263

The paper used in this publication meets the requirements of the American National Standard for Permanence of Paper for Printed Library Materials Z39.48-1984.

10 9 8 7 6 5 4 3 2

To my wife, Adrienne Lehrer

This page intentionally left blank

Contents

Preface to First Edition xiii
Preface to Second Edition xv

1 The Analysis of Knowledge 1
What Is Knowledge? 4
Analysis, 7
The Form and Objectives of an Analysis of Knowledge, 8
The Analysis of Knowledge, 11
Theories of Justification, 15
A Counterexample, 18
Justification Without Falsity: A Fourth Condition, 20
A Final Analysis of Knowledge, 20
Introduction to the Literature, 21
Notes, 21

2 Truth and Acceptance 25
Truth, 26
Acceptance and Knowledge, 32
The Consistency of Knowledge and Belief, 32
Borderline Cases of Knowledge, 36
Memory Without Knowledge, 42
Introduction to the Literature, 43
Notes, 43

3 The Foundation Theory: Infallible Foundationalism 45
Infallible Versus Fallible Foundationalism, 46
Fallible Foundations, 46
The Foundation Theory in General, 48
The Foundation As a Guarantee of Truth, 49
Incorrigible Foundations, 50
Fallibility and Inference: Summary of the Argument, 58
Nomological Infallibility, 59

Meaning and Belief, 59
The Opacity of Incorrigibility, 61
Justification As a Logical Guarantee of Truth for Nonbasic Beliefs, 63
Phenomenalism, 63
Objections to Phenomenalism, 65
Summary, 67
Introduction to the Literature, 68
Notes, 69

4 FALLIBLE FOUNDATIONS 71
Perceptual Belief and Independent Information, 72
Justification and Innocent Belief, 73
Semantics and Justification, 76
Self-Justification and Necessary Truth, 79
Semantic Foundationalism: A Counterexample, 80
Contingent Self-Justification, 83
Probability and Justification: Fallibilistic Foundationalism, 85
Three Concepts of Probability, 86
Probability, Truth, and Basic Belief, 93
Summary: Competence, Success, and Coherence, 94
Introduction to the Literature, 95
Notes, 95

5 THE EXPLANATORY COHERENCE THEORY 97
The Regress, or the Circle, 97
The Traditional Answer: Coherence As Implication, 100
Coherence As Explanation, 101
On the Justification of What Is Explained, 102
Explanatory Coherence and Justification: An Analysis, 103
Objections and Replies to Coherence As Explanation, 108
Self-Explanatory Beliefs, 112
Simplicity and Conservation, 118
Summary, 120
Introduction to the Literature, 121
Notes, 121

6 INTERNAL COHERENCE AND PERSONAL JUSTIFICATION 123
Acceptance and Belief Reconsidered, 124
Justification and Reasonable Acceptance, 125
Justification, Reasonableness, and Coherence, 126
Suggestion of a Unified Theory of Knowledge, 128

Objection Defined, 130
The Justification Game: Replying to a Critic, 132
Answering Objections, 134
The Justification Game and the Definition of Personal Justification, 136
A Foundationalist Objection, 137
The Principle of Trustworthiness of Acceptance, 138
The Virtuous Loop of Reason, 142
The Lottery Paradox, 146
The Advantages of Truth, 147
Introduction to the Literature, 148
Notes, 148

7 COHERENCE, TRUTH, AND UNDEFEATED JUSTIFICATION 151
The Uncharitable Possibility of Error, 151
Undefeated Justification, 153
A Solution: Defeat and the Ultra Justification Game, 160
Truth Connection and the Isolation Objection, 163
Perception, Memory, and Introspection, 165
Knowledge of the Eliminated, 166
A Definition of Undefeated Justification, 168
Knowledge Reduced to Undefeated Justification, 169
Determining Justification, 173
Introduction to the Literature, 174
Notes, 174

8 EXTERNALISM AND THE TRUTH CONNECTION 177
Naturalism, 178
The Advantages of Externalism, 179
The Naturalistic Relation, 182
Objections to Externalism: Information Without Knowledge and the Opacity Objection, 185
Externalism and Justification, 189
Undefeated Justification and Reliabilism, 191
Trustworthiness and Reliability in the Justification Game, 193
Conversion to Irrefutable Justification, 194
Causation and Justification: The Basing Relation, 195
Acceptance, Belief, and Justification, 198
Reliability and the Justification Game, 198
Trustworthiness and Prejudice: An Objection, 200

Externalism, Foundationalism, and Coherence:
An Ecumenical Reconsideration, 201
Introduction to the Literature, 203
Notes, 203

9 SKEPTICISM, VIRTUE, AND CONTEXT 205
Skepticism and Agnoiology, 205
Conception and the Chance of Error, 206
An Answer to Skepticism: Fallibility, Not Ignorance, 208
Intellectual Virtue and Trustworthiness, 209
The Duplication Argument: An Objection, 212
The Merits of Skepticism, 213
Skepticism and Closure: An Externalist Caveat, 214
The Trilemma of Knowledge and Skepticism, 216
Contextualism: Another Solution, 218
Why not Closure? 220
Contextualism and Trustworthiness, 222
The Insufficiency of Reliability and the Role of Explanation, 223
Answering the Skeptic: A Concluding Discourse, 224
Acceptance and Self-Trust: A Reply to the Skeptic, 225
The Virtuous Loop Maximizes Explanation, 228
Summary, 229
Introduction to the Literature, 230
Notes, 231

Bibliography 233
Index 243

Preface to the First Edition

THIS BOOK arose out of my earlier book, *Knowledge,* in more than one way. *Knowledge* had been widely used as a textbook in theory of knowledge courses but then went out of print. Professor Marjorie Clay of Bloomsburg University and Angela Blackburn of Clarendon Press both suggested to me that a textbook based on the previous book would be useful, and Spencer Carr of Westview Press encouraged me with a contract. I am very indebted to these people for their encouragement. The result is what you read.

How similar is this book to *Knowledge?* It is a different book: Chapters 6, 7, and 8 are entirely new; the other chapters contain new material as well but incorporate material from *Knowledge.* In addition, I have included Introduction to the Literature sections that are intended to guide students to further reading on subjects covered in each chapter. Though it has a similar structure to the original, this book was written as a textbook concerning the present state of the art. I have written for students and not for my colleagues, though I hope the latter may find some edification and pleasure in it. I have explained things that I thought students would need explained. I have retained a good deal of the critical argumentation from the earlier book because the same criticisms of various views that seemed cogent to me then seem cogent to me still. I have, however, considered new theories and presented a new form of the coherence theory, leaving the most complicated refinements for my articles. I would encourage those interested in such refinements to read "Knowledge Reconsidered" in *Knowledge and Skepticism* (Westview), edited by Marjorie Clay and myself, as well as my contribution to *The Current State of the Coherence Theory,* edited by John W. Bender.

This book is an attempt to explain foundationalism, the coherence theory, and externalism to students. Of course, I was not bashful about saying where the truth lies. So a form of the coherence theory winds up in the winner's circle, but it is a form that incorporates elements from foundationalism and externalism. I thought it important to advocate a theory rather

than pretend to a balanced presentation of views. Total impartiality is unattainable and the attempt at it soporific. Students, quite understandably, like to feel that what they read is a quest for the truth. The current book is, and it reads that way. I am an analytic philosopher who thinks a philosopher, like others, should attempt to define his or her key terms. Definitions chain together as a result, and it requires some intellectual effort to proceed from beginning to end. However, I have explained each definition with an example, which should make it possible for a student who has difficulty with definitions to grasp the main arguments nonetheless.

I recommend the book to those who like argument and definition turned by examples. Those who are seeking effortless mastery of philosophical profundities will not find that here (nor, I think, anywhere else). The students who like to match wits with argumentation and definition should find a feast here. I wish them a hearty meal with good appetite and encourage them to reject what they find unsavory. They should determine the reason for the offensiveness and prepare a dish of their own—one more to their own liking. To understand philosophy, one must do philosophy. One must seek the truth to know it. That is my advice and, as it turns out, my theory of knowledge as well.

In closing I wish to thank Marian David, Scott Sturgeon, Vann McGee, Gary Gleb, and Jonathan Kvanvig for their critical reflections; my research assistant Barbara Hannan and my editor Spencer Carr for reflections and editorial work; and Lois Day for assisting me in preparing the manuscript. I owe special thanks to my research assistant Leopold Stubenberg for proofreading and compiling the index. I should also like to express my indebtedness to the John Simon Guggenheim Foundation and the National Science Foundation for supporting my research on Thomas Reid, which greatly influenced the current work, and to the National Endowment for the Humanities for sponsoring the Summer Institute in Theory of Knowledge that I directed with Alvin Goldman. This institute, more than any other single factor, was responsible for my writing this book.

Keith Lehrer

Preface to the Second Edition

GRATIFIED BY the reception of the first edition of this book, I undertook a revision to bring it up to date. The reader familiar with the first edition will find that this edition contains new material in Chapter 9 on virtue epistemology, contextualism, and skepticism. Chapters 6–8 are also considerably revised. In Chapter 6 I altered the terminology of "beating competitors" to "answering objections," which seemed more natural and less aggressive. The analysis of undefeated justification in Chapter 7 is based on a simplified and, I think, improved conception of an ultrasystem. I think that the development of the positive theory clarifies the relationship between internalism and externalism—between internal trustworthiness and external truth connectedness. I have tried to clarify and improve both style and theory throughout. There is some overlap between the contents of Chapter 9 and a paper entitled "The Virtue of Knowledge," which is included in a book edited by Abrol Fairweather and Linda Zagzebski to be published by Oxford University Press.

I was greatly assisted in this revision by many people. I am much indebted to my two research assistants, Scott Hendricks and Rachael Poulsen, for working with me on the manuscript and helping develop the coherence theory of knowledge. I am also greatly indebted to a group of students at Simon Fraser University, Stanford University, and Carleton College, where I used the first edition as a text and asked them for critical comment on it, chapter by chapter. They had a profound influence on the second edition, as did my classes and seminars at the University of Arizona. I am especially indebted to J. C. Smith and G. J. Mattey for detailed comments on the first edition and, in the case of Mattey, for comments on the manuscript of the second edition. I am, of course, wholly responsible for the result and for not conforming more closely to excellent suggestions for revision that I have received.

K.L.

This page intentionally left blank

THEORY OF KNOWLEDGE

This page intentionally left blank

1

The Analysis of Knowledge

ALL AGREE THAT KNOWLEDGE is valuable, but agreement about knowledge tends to end there. Philosophers disagree about what knowledge is, about how you get it, and even about whether there is any to be gotten. The question, What is knowledge? is the primary subject of this chapter and of this book. Why approach the theory of knowledge by asking this question? *Epistemology,* the theory of knowledge, and *metaphysics,* the theory of reality, have traditionally competed for the primary role in philosophical inquiry. Sometimes epistemology has won, and sometimes metaphysics, depending on the methodological and substantiative presuppositions of the philosopher.

The epistemologist asks what we know, the metaphysician what is real. Some philosophers have begun with an account of the nature of reality and then appended a theory of knowledge to account for how we know that reality. Plato, for example, reached the metaphysical conclusion that abstract entities, or forms, such as triangularity or justice, are real and all else is mere appearance. He also held that the real is knowable, and he inquired into how we might know this reality.[1] Aristotle, on the contrary, held that individual substances, such as individual statues or animals, are real and inquired as to how we might have knowledge, especially general knowledge, concerning these substances.[2] It is hardly surprising that Plato and Aristotle produced vastly different theories of knowledge when they conceived of the objects of knowledge in such different ways. Their common approach, starting with metaphysics, we might refer to as *metaphysical epistemology.*

The problem with this approach is that the metaphysical epistemologist uncritically assumes that we know the reality posited. He only concerns himself with what such knowledge is like after assuming the nature of real-

ity. This leaves us with the unanswered question of how we know that reality is what the metaphysician affirms it to be and, indeed, begs the question of whether we know such a reality at all.

Other philosophers, most notably René Descartes,[3] turned tables on the metaphysical approach by insisting that we must first decide what we can *know* about what is real and must remain skeptical about what *is* real until we have discovered what we can know. We might refer to this as *skeptical epistemology.*

It seems natural to begin with skepticism with the hope of discovering what we know and what we do not, but if we first pretend to *total* ignorance, we may find no way to remove it. Moreover, we shall lack even the meager compensation of knowing that we are ignorant, for that too is knowledge. Consider, for example, Descartes' attempt to found knowledge on a certain and indubitable premise. To that end he engaged in the project of doubting everything it was possible to doubt, even if the doubt was completely unrealistic or, as he said, hyperbolic. To that end he imagined a powerful demon whose object was to deceive him in all matters that were within the power of demon. The demon might deceive him about abstract matters by confusing his powers of reasoning, just as the demon might deceive him about the objects of the senses by confusing his powers of perception.

The relief from skepticism Descartes obtained was in the claim that he doubted, that he thought, and that he existed. Descartes alleged that he must think and exist to be deceived and, therefore, could not be deceived about the existence of his own thought or about his existence as a thinker. This line of thought has had its detractors. Some of them contend that Descartes was entitled only to the claim that there was a thought, for that was all that was required for the deception to occur. They concluded, therefore, that Descartes was not entitled to the further conclusion that a thinker existed.

The argument appeared to succeed concerning at least the existence of thoughts or ideas, whether or not those thoughts or ideas were true. Moreover, it also succeeds in showing thoughts with a definite content, thoughts that external objects exist, and thoughts of reasoning validly to some conclusion, for only if these thoughts exist could Descartes be deceived into thinking they were true when they were not. Though he might be deceived about the truth of his thoughts, he could not be deceived about their existence. His thoughts and ideas supplied him with a certain and indubitable starting point.

The problem for Descartes and those who followed him in adopting the starting point of ideas, most notably Berkeley[4] and, with an important modification concerning the nature of impressions, David Hume,[5] was to provide some justification for supposing that our thoughts, those arising from our senses, for example, were true and that the objects of our

thoughts, the external world of everyday objects, really did exist. The problem proved intractable given the restriction to a starting point of thoughts and ideas, even if the thoughts included sensory thoughts, sensory appearances or, as Hume called them, impressions. Hume argued convincingly that any attempt to construct an argument from the internal world of ideas and impressions to the external world of objects would require a premise about the correlation between the internal world and external world that was unavailable until one had justified the conclusion that external things existed, which was the very thing to be proved by the argument. When we add Hume's doubt about the power of reasoning for reaching conclusions about the external world to Descartes' demonic doubt, we appear led to skepticism about the external world.

Other philosophers, most recently, extreme materialists,[6] have taken the opposite starting point, beginning with the assumption that we have knowledge of the external material world from observation. Beginning with the assumption that we have knowledge of the external world of matter avoids skepticism concerning the external world, which would appear to be an advantage over beginning with only the premise that we have knowledge of the internal world of ideas.

Unfortunately, this external starting point of matter has a defect exactly analogous to the internal starting point of ideas. If you start with the assumption that you know of the existence of the internal world of ideas from consciousness, you will face the intractable problem of avoiding skepticism concerning the external world of matter. For how can you prove that matter exists assuming only the existence of ideas? Perhaps there is no material world but only a succession of ideas? But if you start with the assumption that you know the existence of matter from observation, you will face the opposite problem of avoiding skepticism concerning the internal world of consciousness. For how can you prove that ideas exist assuming only the existence of matter? Perhaps there are no ideas but only the activation of neurons. Adopting either the internal starting point of ideas or the external starting point of matter leads to skepticism concerning vast domains of knowledge.

Are we then trapped between a method that uncritically assumes our knowledge of reality while assigning priority to metaphysics and one which rejects the assumption that we have knowledge and leads to skepticism? Our approach here will be neither skeptical nor metaphysical. We assign priority to neither metaphysics nor epistemology but attempt to provide a systematic and critical account of prior metaphysical and epistemological assumptions. We refer to this as *critical epistemology.*

We begin with commonsense and scientific assumptions about what is real and what is known. These convictions constitute our data, perhaps even conflicting data if common sense and science conflict. The object of

philosophical inquiry, of which critical epistemology is a fundamental component, is to account for the data. Consider the problems that arise for a philosopher starting with the internal world of consciousness versus those for one starting from the external world of matter. The former finds himself behind an ideal veil of thought trying in vain to reason to an external world while the latter finds himself behind an iron curtain of matter trying in vain to reason to an internal world.

A critical epistemologist will eschew the bias of either starting point for a more balanced and symmetrical point of view, setting out with the premise that we have knowledge of the internal world of our ideas from consciousness and of the external world of matter from observation. Thus, a critical epistemologist, contrary to those who insist on assigning a privileged epistemic status to knowledge of the internal world over knowledge of the external world or vice versa, will insist on a starting point of symmetry between our knowledge of the internal world and the external world and thereby avoid the skeptical conclusions resulting from privileging one kind of knowledge over the other.

The account of the critical epistemologist is essentially and fundamentally critical, however. We are committed as critical epistemologists to reconsider the data of knowledge with which we begin in the light of philosophical investigation. Sometimes we explain the data and sometimes we explain the data away. For the most part, it behooves a critical epistemologist to construct a theory of knowledge explaining how we know the things we think we do, but, in a few instances, a theory may explain why we think we know when we do not. In order to explain what we do know or why we do not, however, we do well to first ask what knowledge is. Indeed, we must do so in order to evaluate the claims of either the metaphysical dogmatist or the epistemological skeptic. It is to this inquiry that we now turn.

What Is Knowledge?

Some have denied that we know what is true or what is false, and they have remained skeptics. Skepticism will have a hearing, but we shall pursue our study as critical epistemologists: We assume people have knowledge. But what sort of knowledge do they have, and what is knowledge anyway? There are many sorts of knowledge, but only one—the knowledge that something is true—will be our concern. Consider the following sentences:

I know the way to Lugano.
I know the expansion of pi to six decimal places.
I know how to play the guitar.
I know the city.
I know John.

I know about Alphonso and Elicia.
I know that the neutrino is electrically neutral.
I know that what you say is true.
I know that the sentence 'some mushrooms are poisonous' is true.

These are but a few samples of different uses of the word 'know' describing different sorts of knowledge.[7] If we are interested in finding out what people have when they have knowledge, we must first sort out the different senses of the word 'know.' Then we may ask our question again, once the word has been disambiguated.

In one sense, 'to know' means to have some special form of competence. Thus, to know the guitar or to know the multiplication tables up to ten is to be competent to play the guitar or to recite the products of any two numbers not exceeding ten. If a person is said to know how to do something, it is this competence sense of 'know' that is usually involved. If I say I know the way to Lugano I mean that I have attained the special kind of competence needed either to get to Lugano or to direct someone there. If I say that I know the expansion of pi expanded to six decimal places, I mean that I have the special competence required to recall or to recite the number pi expanded to six decimal places.[8]

Another sense of 'know' means to be acquainted with something or someone. When I say that I know John, I mean that I am acquainted with John. The sentence 'I know the city' is more difficult to disambiguate. It might mean simply that I am acquainted with the city and hence have the acquaintance sense of 'know,' or it might mean that I have the special form of competence needed to find my way around the city, geographically and/or socially. It also might mean that I know it in both the competence and acquaintance sense of 'know.' This example illustrates the important fact that the senses of 'know' that we are distinguishing are not exclusive; thus, the term 'know' may be used in more than one of these senses in a single utterance.[9]

The third sense of 'know' is that in which 'to know' means to recognize something as true. If I know that the neutrino is electrically neutral, then I recognize something as true, namely, that the neutrino is electrically neutral. When I recognize something as true, I recognize that something is the case, that is, I recognize something as correct information about the world. We may, being careful to use the term 'information' in an ordinary sense of the word, characterize the use of 'know' as the correct information sense of the word. The last three sentences on the list all involve this correct information sense of the word 'know.' It is often affirmed that to know something in the other senses of 'know' entails knowledge in the correct information sense of 'know.' I must have some correct information about Lugano if I know the way to Lugano; about the expansion of pi if I know

the expansion of pi to six decimal places; about the city if I know the city; about the guitar if I know how to play the guitar, and so forth. Thus, the correct information sense of the word 'know' is often implicated in the other senses of the word.

In our study, we shall be concerned with knowledge in the correct *information sense*. The role of such knowledge in human reasoning is essential to its nature. One essential role of knowledge is the employment of it to reason to conclusions, to confirm some hypotheses and refute others. There may be states of mind that are useful to us in a variety of ways (that enable us to avoid harm, for example) but fall short of constituting the sort of knowledge that enables us to reason about what is true and what is false, what is real and what is unreal. States that do not permit us to reason in these ways are not states of human knowledge in the information sense of the word 'know.' It is fundamental to the kind of human knowledge that concerns us in this book that it is inextricably woven into reasoning, justification, confirmation, and refutation. It is required both for the ratiocination of theoretical speculation in science and practical sagacity in everyday life. To do science—to engage in experimental inquiry and scientific ratiocination—one must be able to tell whether one has correct information or not. Scientific knowledge of the world must be based on experimental test and critical evaluation. Engaging in law or commerce requires the same sort of knowledge, which may be used as the premises of critical reflection or claimed as the prizes of informed reasoning.

Correct information is necessary to human knowledge and is useful in picking out the sense of the word 'know' that concerns us, but the possession of correct information is not sufficient for human knowledge in that sense. This sort of knowledge is something beyond the mere possession of information, since one must know that the information one possesses is correct in order to attain knowledge that supplies one with premises for reasoning and the other endeavors described above. It suffices for the mere possession of information, however, that I come to believe something by being informed of it from a trustworthy source. If you tell me something and I believe you, even though I have no idea whether you are a source of truth and correct information about the subject or a propagator of falsehood and deception, I may, if I am fortunate, acquire information when you happen to be informed and honest. This is not, however, knowledge. It is the mere possession of information, which, though necessary for knowledge, is not sufficient for the attainment of knowledge.

Similarly, if I read some gauge or meter and believe the information I receive, though I have no idea whether the instrument is functioning properly, I may thus acquire information, but this is not knowledge. If you doubt this, consider a clock that is not running because it stopped at noon some months ago. As luck would have it, you happen to look at it just at noon

and believe that it is noon as a result. You might, as a result, come to believe it is noon when indeed it is, but that is not knowledge. If the clock is in fact running properly, but, again, you have no idea that this is so, you will have received the information from a reliable source; but your ignorance of the reliability of the source prevents you from recognizing that the information is correct—from knowing that it is correct—even though you may believe it to be so. It is information that we recognize to be correct that yields the characteristically human sort of knowledge that distinguishes us as adult cognizers from machines, other animals, and even our infant selves.

Some philosophers, choosing to place emphasis on the similarity between ourselves and these other beings, may insist that they have knowledge when they receive information in some technical sense of the word.[10] This is a verbal dispute in which we shall not engage, for it is profitless to do so. We shall remain content with the observation that our most cherished scientific achievements, the discovery of the double helix, for example, and our most worthy practical attainments, the development of a system of justice, for example, depend on a more significant kind of knowledge. This kind of knowledge rests on our capacity to distinguish truth from error and to make use of it to reason about what is true and what is false.

Analysis

To indicate the information sense of the word 'know' as being the one in question is quite different from analyzing the kind of knowledge we have picked out. What is an analysis of knowledge? An analysis is always relative to some objective. It does not make any sense simply to demand the analysis of goodness, knowledge, beauty, or truth without some indication of what purpose such an analysis is supposed to achieve. To demand the analysis of knowledge without specifying further what you hope to accomplish with it is like demanding blueprints without saying what you hope to build. Before asking for such an analysis, we should explain what goals we hope to achieve with it.

First, let us consider the distinction between analyzing the meaning of the term 'know' and analyzing the kind of knowledge denoted. Many philosophers have been interested in the task of analyzing the meaning of the word 'know.'[11] Indeed, many would argue that there is no need for philosophical analysis once we have a satisfactory analysis of the meaning of the term 'know.' This restrictive conception of philosophical analysis is sustained by a dilemma: either a theory of knowledge is a theory about the meaning of the word 'know' and semantically related epistemic terms or it is a theory about how people come to know what they do. The latter is not part of philosophy at all, but rather that part of psychology called learning theory.

It follows that if a theory of knowledge is part of philosophy, then it is a theory about the meaning of the word 'know.' That is the argument, and it is one that would reduce the theory of knowledge to a theory of semantics.

It is not difficult to slip between the horns of the dilemma. A theory of knowledge need not be a theory about the meaning of epistemic words any more than it need be a theory about how people come to know what they do. Instead, it may be a theory of what conditions must be satisfied and how they may be satisfied in order for a person to know something. When we specify those conditions and explain how they are satisfied, then we have a theory of knowledge. An analogy should be helpful at this point. Suppose a person says that there are only two kinds of theories about physical mass. Either a theory of matter is a theory about the meaning of 'mass' and semantically related physical terms or it is a theory about how something comes to have mass. This dichotomy would be rejected on the grounds that it leaves out the critical question of what mass is or, to put it another way, it leaves out the question of what conditions must be satisfied for something to have a given mass. A theoretician in physics might be concerned with precisely the question of what conditions are necessary and sufficient for an object to have mass or, more precisely, to have a mass of m. The answer, of course, is $m = a/f$. Similarly, a philosopher might be concerned with precisely the question of what conditions are necessary and sufficient for a person to have knowledge or, more precisely, to know that p.

Some philosophers might question whether it is possible to give necessary and sufficient conditions for knowledge, but the finest monuments of scientific achievement mark the refutation of claims of impossibility. Obviously, a necessary and sufficient condition for the application of the expression 'S knows that p' is precisely the condition of S knowing that p. This could be made less trivial with little difficulty. The objection to the idea that a philosopher can discover necessary and sufficient conditions for knowledge may rest on the confused idea that a set of conditions necessary and sufficient for the application of a term constitutes a kind of recipe for applying terms that would enable us to decide quite mechanically whether the term applies in each instance. However, we may, without taking any position on the question of whether such a recipe can be found for applying the term 'know,' state flatly that this is not the purpose of our theory of knowledge or the analysis of knowledge incorporated therein. Our interests lie elsewhere.

The Form and Objectives of an Analysis of Knowledge

We shall then approach the question, What is knowledge? with the objectives of formulating necessary and sufficient conditions for a person having

knowledge (in the information sense of the term 'know') and of explaining how those conditions may be satisfied. Our project is contiguous with scientific investigations having analogous objectives. Our conception of analysis is indebted to both Rudolf Carnap and W. V. O. Quine.[12] Carnap proposed that philosophy should aim at *explication*. This is a kind of analysis aiming at the generation of philosophically and scientifically useful concepts. More specifically, explication aims at producing concepts useful for articulating laws and theories. For example, the explication of 'fish' so as to exclude whales from the class of fish generates a scientifically useful concept for the purpose of formulating laws. One such law is that fish are cold-blooded, to which whales would constitute a counterinstance if whales were included in the class of fish. When, however, we take this purpose of explication seriously and adopt the strategy of providing analyses of this sort in philosophy, then, as Quine argued, there can be no clear boundary between philosophy and science. Our reasoning is that it is surely the purpose of science as well as philosophy to provide concepts to facilitate the formulation of laws and theories.

Thus, we contend that the distinction between philosophy and theoretical science is a bogus distinction, whether viewed historically or systemically.[13] Historically, it is clear that the special sciences break off from philosophy when some theory emerges that deals with a circumscribed subject in a precise and satisfactory manner. Philosophy remains the residual pot of unsolved intellectual problems. To date, theories of knowledge have remained in the pot. We do not claim that the current study or other recent research has brought us to the point where the theory of knowledge should be poured out into a special science, but we hope that we are closer to that goal than some suspect and others fear.

A formulation of an analysis of knowledge may be expressed by an equivalence. Again, the analogy with mass is helpful. An analysis of mass may be given in an equivalence of the following form:

O has a mass of *n* if and only if . . .

where the blank to the right of the equivalence is filled with a sentence describing a set of necessary and sufficient conditions. Similarly, an analysis of knowledge may be given in an equivalence of the following form:

S knows that *p* if and only if . . .

where the blank to the right of the equivalence is filled with a sentence describing a set of necessary and sufficient conditions for knowledge.

When considering candidates for such sets of conditions, we ask whether there is any counterexample to the proposed analysis. What is a

counterexample? First of all, any experiment of fact or thought that would falsify the resulting equivalence is a counterexample. To say that there is no experiment of thought to falsify the equivalence means that we can think of no logical possibility which is consistent with other postulates of the theory under consideration that would result in one side of the equivalence being satisfied and the other not. We shall begin by considering any logically possible case as a potential counterexample to a theory of knowledge. We may decide eventually, however, that some examples, though logically possible, are so remote in terms of real possibility that they do not constitute realistic objections to an analysis of actual human knowledge.

In addition to being immune from counterexamples, such an equivalence will be a suitable analysis only if it facilitates reaching our epistemic objectives. Thus, though some analyses are definitely mistaken because we can find acceptable counterexamples, there are other equivalences that fail to constitute satisfactory analyses simply because they are unenlightening. To say that a person knows that *p* if and only if it is known to the person that *p*, though this is immune from counterexamples, would completely fail to explain or inform. The explanatory role of an analysis is of fundamental importance and must be appealed to in support of an analysis. It is important, therefore, to consider at the outset what sort of enlightenment one is seeking, that is, what one is attempting to explain by means of an analysis. We shall be concerned with an analysis that will be useful for explaining how people know that the input (the reports and representations) they receive from other people, their own senses, and reason is correct information rather than error and misinformation. A person may receive a representation that *p* as input without knowing that the representation is correct and, therefore, without knowing that *p*.

Suppose, for example, that some person unknown to me tells me that all the perch in the Genesee River will be killed by a pollutant that has raised the temperature of the water two degrees. I might believe what I am told, being gullible, but I do not know whether my informant knows whereof she speaks. Consequently, I do not know the perch will die. My informant may be knowledgeable. I may possess accurate information as a result of believing what I was told, but I do not know that the report is correct. Similarly, if I possess some information in memory but no longer know whether it is correct information, whether it is something I accurately remember or just something I imagine, I am again ignorant of the matter. If, on the other hand, I know that the information I possess is correct, then I have knowledge in the requisite sense.

One test of whether I know that the information I possess is correct is whether I can answer the question of how I know that the information is correct or how I would justify claiming to know. Such questions and the

answers provided are the basis for critical discussion and rational confrontation in scientific inquiry and everyday life. If I know that some information is correct, I will be in a position to use it as a premise in reasoning to confirm or refute some claim or other. If I do not know the information to be correct, then, even if it is correct, it does not supply me with a premise nor can I reply to relevant queries including, most saliently, how I know. The replies to such queries show us whether or not the conditions for knowledge have been satisfied. If a person claims to know something, how well she answers the question, How do you know? will determine whether we accept her claim. Consequently, our analysis of knowledge should explain how a person knows that her information is correct and, so that her knowledge may play its essential role in reasoning, how her knowledge claims are justified.

The foregoing remarks indicate why we shall not be concerned with the sort of knowledge attributed to animals, small children, and simple machines that store information, such as telephones that store telephone numbers. Such animals, children, or machines may possess information and even communicate it to others, but they do not know that the information they possess is correct nor are they in a position to use the information in the requisite forms of ratiocination. They lack any conception of the distinction between veracity and correct information, on the one hand, and deception and misinformation, on the other. Any child, animal, or machine that not only possesses information but knows whether the information is correct is, of course, a candidate for being a knowing subject. In those cases in which such knowledge is lacking, however, we shall assume ignorance in the information sense of knowledge under investigation here.

The Analysis of Knowledge

With these initial remarks to guide us, we shall now offer a preliminary analysis of knowledge. Each condition proposed will be the subject of subsequent chapters. Moreover, in the case of some controversial conditions, we shall not undertake a detailed defense in the present chapter. Our intention here is only to provide the analysis with some intuitive justification that will subsequently be developed and defended.

A Truth Condition

The first condition of knowledge is that of *truth*. If I know that the next person to be elected president of the United States will have assets of at least $1 million, then it must be true that the next president will have assets of at least $1 million. Moreover, if the next person to be elected president will, in fact, not have assets of at least $1 million, then I do not know the

next president will have assets of at least $1 million. If I claim to know, my knowledge claim is incorrect. I did not know what I said I did. Thus, we shall accept the following conditionals:

(iT) if *S* knows that *p*, then it is true that *p*

and

(iT') If *S* knows that *p*, then *p*.

The two conditionals are equivalent for all those cases in which instances of the following principle, which articulates the absolute theory of truth to be discussed in the next chapter, are necessarily true:

(AT) It is true that *p* if and only if *p*.

It is true that the United States has a president if and only if the United States has a president, and this is necessarily true. The equivalence of the conditionals

If Lehrer knows that the United States has a president, then it is true that the United States has a president.

and

If Lehrer knows that the United States has a president, then the United States has a president.

is a result of the necessary truth of

It is true that the United States has a president if and only if the United States has a president.

We shall find in the next chapter, however, that in spite of the innocent and even trivial appearance of (AT), the absolute theory of truth, it leads to paradox in some instances.

An Acceptance Condition

The second condition of knowledge is *acceptance.* If I deceitfully claim to know that Jan and Jay married on 31 December 1969, when I do not accept it, then I do not know Jan and Jay were married on that date even if they were married then. If I do not accept that *p*, then I do not know that *p*. Thus, the following conditional expresses a condition of knowledge:

(iA) If *S* knows that *p*, then *S* accepts that *p*.

A more familiar and quite similar condition would require belief as a condition of knowledge as follows:

(iB) If *S* knows that *p*, then *S* believes that *p*.

These two conditions would be equivalent if the following equivalence were necessarily true:

(AB) *S* accepts that *p* if and only if *S* believes that *p*.

Principle (AB) is not true, however. Acceptance is an attitude defined in terms of some purpose, that is, the expression to *accept that p* is an abbreviation or shorthand for the more explicit expression to *accept that p for O*, where *O* is some specific objective or purpose. It involves an evaluation of whether the attitude fulfills the purpose. Acceptance requisite to knowledge is a special kind of acceptance for an intellectual purpose concerned with truth. It is accepting something for the *epistemic* purpose of attaining truth and avoiding error with respect to the very thing one accepts. More precisely, the purpose is to accept that *p* if and only if *p*.

Belief may result from the pursuit of some purpose, but it is not *defined* in terms of any purpose. Belief may, for example, serve the epistemic purpose of attaining truth and avoiding error, but it is not defined in terms of that purpose in the way that acceptance is. Sometimes, moreover, we believe things that we do not accept for this epistemic purpose. We may believe something for the sake of felicity rather than from a regard for truth. We may believe that a loved one is safe because of the comfort of believing this, though there is no evidence to justify accepting this out of regard for truth, indeed, even when there is evidence against it. So, there are cases in which we do not accept something for the purposes of attaining truth and avoiding error but we believe it, nonetheless. It is the acceptance of something defined in terms of the epistemic purpose of attaining truth and avoiding error rather than mere belief that is the required condition of knowledge.

Some philosophers have insisted that a person may know something is true even though she lacks conviction of its truth. Others, in diametric opposition, have contended that a person only knows that something is true when she is sure, or certain, of the truth of what she believes. Thus, some philosophers have denied condition (iB) on the grounds that a person may know something to be true that she does not believe at all,[14] and others have maintained that for a person to know something to be true she must believe it to be true with considerable certainty.[15] Our proposal is that ac-

ceptance aiming at the epistemic purpose rather than belief, condition (iA) rather than (iB), is what is needed. A person need not have a strong feeling or conviction that something is true in order to know that it is. What is required is acceptance of the appropriate kind, acceptance in the interest of obtaining a truth and avoiding an error in what one accepts.

Acceptance of other kinds, having other purposes, must also be acknowledged. For example, a person may accept something for the sake of argument, for the purpose of testing it as a hypothesis, or to make somebody happy. It is, however, a special kind of acceptance, namely, acceptance aimed at attaining truth and avoiding error that is a necessary condition of knowledge. Other kinds of acceptance are not necessary for knowledge.

The difference between acceptance of the sort required for knowledge and mere belief consists in the fact that acceptance involves evaluation in terms of the epistemic purpose. Belief may fulfill some purpose without any evaluation in terms of the purpose. My belief that *p* may fulfill the epistemic purpose if *p* is true, even though I have not evaluated the matter in terms of that objective but have only sought to conform to authority to obtain peace of mind. Nevertheless, acceptance that *p* involving evaluation in terms of the epistemic purpose may coincide with belief that *p*. We may, therefore, expect the appropriate kind of acceptance to be accompanied by a kind of belief, but we should not assume that belief is the same as acceptance.

We gain some continuity with tradition as well as some expository simplification by speaking as though acceptance was always accompanied by belief. When it is, we may, consequently, speak of belief as a condition of knowledge for the sake of tradition, but we shall recall that it is acceptance aimed at truth that is genuinely required for knowledge, not the belief that accompanies it, and use the terms "accept" and "acceptance" when precision is needed.

A Justification Condition

Accepting something that is true does not suffice for knowledge. If I accept something without evidence or justification, that my wife has exactly fourteen dollars in her purse, for example, and, as luck would have it, this turns out to be right, I fall short of *knowing* that what I have accepted is true. Thus, we require a third condition affirming the need for *justification.* While we allowed that a person need not be completely certain of *p* in order to know that *p,* we shall insist that he be justified in his acceptance of *p* in order to be said to know that *p*. Moreover, the person must be justified in way that would justify him in accepting that he knows, if he considers whether he does.

The reason for requiring that a person be justified in this way, rather than, for example, simply being reasonable to accept what he does, is to re-

quire that the justification is strong enough for the purpose of analyzing knowledge. Reasonableness may be too weak a requirement. I may be reasonable in accepting that my secretary is in her office now because she is ordinarily there at this time. Not being there myself, however, I do not know that she is there. Though I am reasonable in accepting that she is there, I am not justified in a way which would allow me to say that I know she is. I am unable to exclude the possibility that she is out of the office on an errand, for example, and, in that way, my reasonableness falls short of justification. Our condition may be formulated as follows:

(iJ) If *S* knows that *p,* then *S* is justified in accepting that *p.*[16]

Sometimes when a speaker says another person is justified in accepting something, the speaker may say this because he, the speaker, has adequate evidence, without implying that the other person does. If you say "Alice thinks she is broke" because she has made a miscalculation in checking her account, I might reply, having emptied her account electronically without her knowledge, "She is justified in thinking that," using the expression "justified" to refer to the information I possess but she lacks about her account. We shall not use the expression "justified" in the way I use it in this example. When we say that *S* is justified, we shall mean that her acceptance is based on adequate evidence of hers, that is, that she is justified by the evidence she has in accepting that *p*. Thus, that I am justified in accepting that *p* by the evidence I have does not by itself warrant my saying that another is justified in her acceptance of *p*. She too must have evidence that justifies her acceptance before she is, in the required sense, justified in accepting that *p*. The moral of the preceding remarks is that we shall not be enslaved to ordinary thought and speech when we speak of "justification" but, for the sake of theoretical advantage, we shall delete unwanted implications and allow expedient expansion within the theory of justification articulated below.

Theories of Justification

There are three kinds of theories of justification that we shall discuss in detail in subsequent chapters. These theories constitute the heart of a theory of knowledge. The first kind of theory is a *foundation* theory of justification. According to foundationalists, knowledge and justification are based on some sort of foundation—the first premises of justification. These premises provide us with basic beliefs that are justified in themselves, or self-justified beliefs, upon which the justification for all other beliefs rests.[17]

The motives for such a theory are easy to appreciate. If one thinks of justification in terms of an argument for a conclusion, it appears that justifica-

tion must either continue infinitely from premise to premise, which would be an infinite regress, or argumentation must end with some first premises. Suppose I accept something and, indeed, claim to know it. You ask how I know it. I appeal to something else I accept and claim to know. You ask how I know that. Each time I attempt to justify what I accept by something else I accept. At some point, it would seem that I would have to end this regress of justification, admitting I have arrived at a first premise I can justify no further, to avoid continuing infinitely. Such first premises would be basic beliefs justified without appeal to other premises. This alternative is the one chosen by the foundation theory.

Basic beliefs constitute the evidence in terms of which all other beliefs are justified, according to the foundation theorist. Some empiricist philosophers affirm the existence of basic beliefs concerning perception (*I see something red,* for example) or more cautious beliefs about mere appearance (*I am appeared to in a reddish way,* for example) and maintain that all justification would be impossible without them. They aver that unless there are some basic beliefs to which we may appeal in justification, we lack a necessary starting point and fall victim to skepticism. In the absence of basic beliefs the whole edifice of justification would collapse for want of a foundation.

Not all epistemologists agree with this contention. A second kind of theory of justification, a *coherence* theory, denies the need for basic beliefs. Coherentists argue that justification must be distinguished from argumentation and reasoning. For them, there need not be any basic beliefs because all beliefs may be justified by their relation to others by mutual support.[18] The edifice of justification stands because of the way in which the parts fit together and mutually support one another rather than because they rest on a concrete foundation of basic beliefs.

How can a coherence theory of justification avoid an unceasing regress proceeding from premise to premise without appeal to basic beliefs? We may avoid a regress of justification without appeal to basic beliefs, says the coherence theorist, because beliefs are justified by the way they agree or cohere with a system of beliefs. Such coherence is a form of mutual support that need not require argumentation; consequently, the regress, which is a regress of argumentation, is avoided because it is not needed for the justification of the belief resulting from coherence with a background system. My perceptual belief that I see something red, for example, is justified because of the way it coheres with a system of beliefs that tells me under what conditions I can tell something red when I see it. It is coherence rather than reasoning or argumentation that yields justification.

One typical objection raised against a coherence theory by a foundationalist is that justification arising from coherence is circular. The foundationalist objection is that reasoning which uses as a premise the very conclusion

to be proven is viscously circular and proves nothing. Circular reasoning, the foundationalist objects, though it may be completed, is no more effective than regressive reasoning to justify a conclusion. The coherence theorist replies that coherence is a form of mutual support among the things a person accepts, which need not take the form of argumentation or reasoning. The coherence theorist must, however, provide some explanation of why remaining within the circle of acceptances is compatible with the idea that coherence yields justification.

This dispute between the foundation theorist and the coherence theorist is joined by a third party, the defender of an *externalist* theory, who disagrees with both parties to the dispute. We need neither basic beliefs nor coherence to obtain knowledge, the externalist contends, but rather the right sort of external connection between belief and reality to obtain knowledge.[19] Causality is one contender for the role of the needed external connection. What makes my belief that I see something red a case of knowledge on such an account is that my belief is *caused* by my seeing some external red object. Such philosophers may even go so far as to deny that justification is necessary for knowledge, contending that only the desired external connection is necessary. We may, however, do the externalist no injury by looking upon the external connection as providing us with a kind of external justification.

The foundation theorist and the coherence theorist may together protest, however, that a person totally ignorant of the external relationship of her belief, the causal history of her belief, for example, will not know that her belief is true unless it is justified by basic beliefs or coheres with a system of beliefs. The externalist will reply that the appropriate external connection requires neither basic beliefs nor coherence to yield knowledge. We leave the dispute unresolved here to become the centerpiece of our inquiry later.

We shall eventually argue, however, that justification is justification based on coherence within an acceptance system of a person, which converts into knowledge with the addition of some features adapted from the foundation theorist and the externalist. From the former, we shall take the insight that some beliefs are justified without being conclusions of argumentation; from the latter, we shall incorporate the idea that a system yielding coherence may contain correct representations of how our beliefs are connected to reality. We shall find that the engine of justification is what a person accepts in the quest for truth combined with the success of the quest.

Most philosophers have thought that knowledge must be based on some objective method for assessing claims of truth or falsity. Some thought the test was that of experience, others of reason, and there have been mixed methodologies as well. All have assumed that acceptance must be checked

in some objective manner. They have repudiated with epistemic horror the idea that acceptance of any sort could convert into the sort of justification required for knowledge. That a person accepts something for whatever purpose is far too subjective a datum to serve as a solid basis for such justification. Even philosophers who argue that some beliefs are self-justified have sought some principle by means of which we can determine which beliefs are self-justified and which not. They have held, too, that we must somehow transcend the subjectivity of acceptance in order to demarcate the area of justification. This conception has become so ingrained philosophically as to impose itself on common sense. However, the assumption that there is some objective method for distinguishing the honest coin of justified acceptance from the counterfeit of the unwarranted shall not go unexamined. We shall study in some detail theories that rest on this assumption but, to warn the reader fairly in advance, no such theory shall prevail once we have exhibited our mint for epistemic approval.

The theory of justification we shall ultimately defend may strike some as closely aligned with skepticism. We shall examine this charge, but even here it should be noted that our sympathies with the writings of the philosophical skeptics of the past are strong. Too often contemporary writers seek the most effective method for liquidating the skeptic without asking whether his teaching may not be of more importance than his mode of burial. Since the most brilliant philosophers of past and present have been skeptics of one form or another, it would behoove those who study skepticism to consider whether these skeptics have some truth in their grasp. We claim they do. At the heart of the skeptic's position is the insight of our fallibility in what we accept. Nevertheless, though we are fallible in what we accept, when acceptance has the purpose of obtaining truth and avoiding error, it may achieve these objectives in a way that yields justification and knowledge. When we strike the right match between the way of acceptance and the way of truth, we obtain the illumination of knowledge.

A Counterexample

Some philosophers have suggested that the conditions which we have considered necessary for knowledge are jointly sufficient for knowledge as well.[20] This would amount to affirming the following equivalence as an analysis of knowledge:

> *S* knows that *p* if and only if it is true that *p*, *S* accepts that *p*, and *S* is justified in accepting that *p*.

In short, knowledge is justified true acceptance. Nevertheless, this analysis has been forcefully disputed and requires amendment.[21]

Edmund Gettier has presented us with counterexamples to the claim that knowledge is justified true acceptance, for example: Suppose a teacher wonders whether any member of her class owns a Ferrari and, moreover, suppose that she has very strong evidence that one student, a Mr. Nogot, owns a Ferrari. Mr. Nogot says he does, drives one, has papers stating he does, and so forth. The teacher has no other evidence that anyone else in her class owns a Ferrari. From the premise that Mr. Nogot owns a Ferrari, she draws the conclusion that at least one person in her class owns a Ferrari. The woman might thus be justified in accepting that Mr. Nogot owns a Ferrari.

Now imagine that, in fact, Mr. Nogot, evidence to the contrary notwithstanding, simply does not own the Ferrari. He was out to deceive his teacher and friends to improve his social status. However, another student in the class, a Mr. Havit, does own a Ferrari, though the teacher has no evidence or inkling of this. In that case, the teacher would be quite correct in her belief that at least one person in her class owns a Ferrari, only it would not be Mr. Nogot, who she thinks owns one, but Mr. Havit, who owns one though she does not think he does, instead. In this case, the teacher would have a justified true belief when she accepts that at least one person in her class owns a Ferrari, but she could not be said to know that this is true because it is more due to good fortune than good justification that she is correct.[22]

To put the argument schematically, Gettier argues that a person might be justified in accepting that *F* by her evidence, where *F* is some false statement, and deduce *T* from *F,* where *T* is some true statement. Having deduced *T* from *F,* which she was justified in accepting, though it was false, the person would then be justified in accepting that *T.* Assuming she accepts that *T,* it would follow from the analysis that she knows that *T.* In such a case, the belief that *T* will be true, but the justification the person has for accepting *T* to be true depends on the reasoning of *T* from *F.* Since *F* is false, it is a matter of luck that she is correct in her belief that *T.*[23] The argument depends on the assumption that a person can be justified in accepting *F* when *F* is false, which assumes our fallibility and the fallibility of our justification.

One might be inclined to reply that reasoning from a false statement can never yield justification, but similar examples may be found that do not seem to involve any reasoning. An example taken from R. M. Chisholm illustrates this. Suppose a man looks into a field and spots what he takes to be a sheep.[24] The object is not too distant and the man generally knows a sheep when he sees one. In such a case, it would be natural to regard the man as being justified in accepting that he sees a sheep in the field without any reasoning at all. Now imagine that the object he takes to be a sheep is not a sheep but a dog. Thus, he does not know that he sees a sheep. Imag-

ine, further, that an object in the deeper distance, which he also sees but does not think is a sheep, happens in fact to be a sheep. So it is true that the man sees a sheep and, moreover, accepts and is justified in accepting that he sees a sheep. Of course, he still does not know that he sees a sheep because what he takes to be a sheep is not, and the sheep that he sees he does not take to be a sheep.

Justification Without Falsity: A Fourth Condition

In the two cases we have described, a person has justified true acceptance but lacks knowledge and in one case does not reason to what he thus accepts from any false statement. There is some merit, however, in the idea that the falsity of some statement accounts for the lack of knowledge. Somehow, it is the falsity of the two statements (that Mr. Nogot owns a Ferrari and that what the man takes to be a sheep really is one) that accounts for the problem. It is false that Mr. Nogot owns a Ferrari, and it is also false that what the man takes to be a sheep is really a sheep (because it is a dog). We may say that in the first case the teacher's justification for her belief—at least one person in her class owns a Ferrari—depends on the false statement that Mr. Nogot owns a Ferrari and that in the second case the man's justification for his belief that there is a sheep in the field depends on the false statement that what he takes to be a sheep is really a sheep.

We shall explore the kind of dependence involved subsequently, but here we may notice that the teacher would be unable to justify her acceptance that there is a Ferrari owner among her students were she to concede the falsity of the statement that Mr. Nogot owns a Ferrari. Similarly, the man would be unable to justify his acceptance that there is a sheep in the field were he to concede the falsity of the statement that what he takes to be a sheep really is a sheep.

To render our analysis impervious to such counterexamples, we must add the condition that the justification that a person has for what she accepts must not depend on any false statement—whether or not it is a premise in reasoning. We may thus add the following condition to our analysis:

> (iD) If *S* knows that *p*, then *S* is justified in accepting that *p* in some way that does not depend on any false statement.[25]

A Final Analysis of Knowledge

The preceding condition enables us to complete our preliminary analysis of knowledge as follows:

(AK) *S* knows that *p* if and only if (i) it is true that *p*, (ii) *S* accepts that *p*, (iii) *S* is justified in accepting that *p*, and (iv) *S* is justified in accepting *p* in some way that does not depend on any false statement.

Our next task is to examine each of these conditions of knowledge in order to formulate a theory of knowledge explaining how and why claims to knowledge are justified. We begin in the next chapter with an account of truth and acceptance and then proceed to consider theories of justification. The discussion of such theories will lead us to an account that brings central features of the various theories under the umbrella of a coherence theory. The correct theory of knowledge must provide the correct blend of acceptance and truth in what is accepted, that is, the right match between mind and reality. A match between mind and world sufficient to yield knowledge rests on coherence with a system of things we accept, our acceptance system, which must include an account of how we may succeed in our quest for truth. This account of how we succeed must be undefeated and irrefutable by errors in what we accept to convert our justification to knowledge. When we have a theory that satisfies these requirements before us, we shall return, at the end, to the speculations of skeptical and metaphysical epistemologists supplied with the scale of knowledge to weigh their claims.

Introduction to the Literature

There are a number of good introductions to the theory of knowledge. Two general collections of essays pertaining to both classical and contemporary literature are *Human Knowledge,* edited by Paul K. Moser and Arnold Vander Nat; and *The Theory of Knowledge: Classic and Contemporary Readings*, edited by Louis P. Pojman. A traditional collection still worth consulting is *Essays on Knowledge and Justification,* edited by George S. Pappas and Marshall Swain. Two splendid and readable traditional introductions are *The Problems of Philosophy,* by Bertrand Russell; and *The Problem of Knowledge,* by Alfred J. Ayer. There are some excellent recent textbooks written by single authors. The best are *Contemporary Theories of Knowledge,* by John L. Pollock and Joseph Cruz; *Belief, Justification, and Knowledge,* by Robert Audi; *Theory of Knowledge,* 3d ed., by Roderick Chisholm; and *An Introduction to Epistemology,* by Jack S. Crumley II.

Notes

1. Plato *Symposium; Phaedo.*
2. Aristotle *Metaphysics,* Z.
3. Descartes *Meditations,* I.

4. George Berkeley, *A Treatise Concerning the Principles of Human Knowledge*, ed. Colin M. Turbayne (Indianapolis: Bobbs-Merrill, 1957).

5. David Hume, *A Treatise of Human Nature* (Oxford: Clarendon, 1888) [originally published in 1739]; and *An Inquiry Concerning Human Understanding* (Indianapolis: Bobbs-Merrill, 1977), XII, pt. 1.

6. Cf. Paul Churchland, *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain* (Cambridge: MIT Press, 1995).

7. For an account of knowledge intended to unify these conceptions, see Colin McGinn, "The Concept of Knowledge," *Midwest Studies in Philosophy* 9 (1984): 529–54.

8. See John Hartland-Swann, *An Analysis of Knowing* (London: Allen & Unwin, 1958), chap. 4; Gilbert Ryle, "Knowing How and Knowing That," *Proceedings of the Aristotelian Society* 46 (1945–46): 1–16.

9. Bertrand Russell uses the expression "knowledge by acquaintance" but in a somewhat more technical sense. See his "Knowledge by Acquaintance and Knowledge by Description," *Proceedings of the Aristotle Society* 11 (1910–11): 108–28; reprinted with some alterations as chapter 5 in *The Problems of Philosophy* (London: Oxford University Press, 1959), 46–59.

10. Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981).

11. For example, see A. J. Ayer's *The Foundations of Empirical Knowledge* (New York: St. Martin's, 1955); and *The Problem of Knowledge* (Harmondsworth, U.K.: Penguin, 1957).

12. Rudolf Carnap, introduction to *The Logical Foundations of Probability* (Chicago: University of Chicago Press, 1962); W. V. O. Quine, "Two Dogmas of Empiricism," in *From a Logical Point of View* (Cambridge: Harvard University Press, 1953), 20–46.

13. The attack on this distinction is due, most recently and impressively, to Quine's discussion in "Two Dogmas of Empiricism." The theme is developed further in his later works, e.g., "Epistemology Naturalized," in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969). Quine concludes with the proposal that we give up traditional epistemology based on a theory of justified belief. His argument does not show that we cannot distinguish justified from unjustified belief, however.

14. Colin Radford, "Knowledge—By Examples," *Analysis* 27 (1966): 1–11.

15. See G. E. Moore's "Certainty," in *Philosophical Papers* (London: Allen & Unwin, 1959), 226–51; and A. J. Ayer's *The Problem of Knowledge*, chap. 1, sec. 3, 14–26.

16. Ayer, *Problem of Knowledge*, 31–35, formulates the condition as the right to be sure. Chisholm formulates it as having adequate evidence in *Perceiving: A Philosophical Study* (Ithaca, N.Y.: Cornell University Press, 1957), 5, 17, and as something being evident for a man in *Theory of Knowledge* (Englewood Cliffs, N.J.: Prentice-Hall, 1966), 18–23. This reference is to the first edition.

17. Foundation theories are defended in John L. Pollock and Joseph Cruz, *Contemporary Theories of Knowledge* (Lanham, Md.: Rowman & Littlefield, 1999), chap. 5; R. M. Chisholm, *The Foundations of Knowing* (Minneapolis: University of Minnesota Press, 1982); and Paul Moser, *Empirical Justification* (Dordrecht: Rei-

del, 1985). Pollock's theory is unusual in that the foundational states are not beliefs; he calls his theory a "non-doxastic" version of "direct realism." Some might not call such a theory a foundation theory at all, but it is more closely allied with a foundation theory than with any other type of epistemological theory.

18. Coherence theories are defended in Keith Lehrer, *Knowledge* (Oxford: Clarendon, 1974); and Laurence BonJour, *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press, 1985).

19. Representative externalist theories can be found in Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981); Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981); and Alvin Goldman, *Epistemology and Cognition* (Cambridge: Harvard University Press, 1986).

20. Ayer and Chisholm defend similar analyses. See Ayer, *Problem of Knowledge;* and Ayer, *Foundations of Empirical Knowledge;* Chisholm, *Perceiving.*

21. See Edmund Gettier Jr., "Is Justified True Belief Knowledge?" *Analysis* 23 (1963): 121–23. Russell made a similar observation in *Problems of Philosophy,* 132.

22. These examples and related ones are taken from Keith Lehrer, "Knowledge, Truth, and Evidence," *Analysis* 25 (1965): 168–75. This article and others on the same topic are included in Michael Roth and Leon Galis, eds., *Knowing: Essays in the Theory of Knowledge* (New York: Random House, 1970).

23. Gettier, "Is Justified True Belief Knowledge?"

24. The sheep example comes from R. M. Chisholm, *Theory of Knowledge,* 23 n. 22.

25. This proposal is similar to one made in the article by Lehrer cited above as well as by others in the series of articles elicited by the Gettier example.

This page intentionally left blank

2

Truth and Acceptance

WE HAVE SAID that knowledge implies acceptance and truth, which commits us to the following two implications:

(iA) If *S* knows that *p*, then *S* accepts that *p*

and

(iT) If *S* knows that *p*, then it is true that *p*.

In this chapter we shall discuss and clarify the conceptions of acceptance and truth that are necessary conditions of knowledge.

These two conditions of knowledge are closely connected because the kind of acceptance necessary for knowledge is directed at obtaining truth and avoiding error. It is important to recall that a person may accept something for some purpose other than a concern for truth. For example, a *fideist,* someone whose faith is not based on reason, might accept that God exists for the sake of piety without any concern for evidence concerning whether it is true that God exists. Sometimes a person accepts something for the sake of piety or felicity without any concern for the truth of what she thus accepts. That, however, is not the sort of acceptance that is a condition of knowledge. On the contrary, the sort of acceptance requisite to knowledge is precisely acceptance concerned with obtaining truth and avoiding error with respect to the very thing accepted.

I know that the structure of the human genome is a double helix. This is something that I accept in the interests of obtaining truth and avoiding error. The sort of acceptance that is a condition of knowledge is accep-

tance aimed at truth, and it is in this way that acceptance and truth are connected. Can we give a more precise account of what is meant by saying that a person accepts something in the interests of obtaining truth and avoiding error with respect to the thing accepted? It is important to notice, first of all, that these two interests—obtaining truth and avoiding error—are not in harmony with each other. If I were interested simply in avoiding errors in what I accept, the rational strategy would be to accept nothing; for, if I were to accept nothing, then I would accept nothing false. On the other hand, if I were interested simply in accepting truths, the rational strategy would be to accept everything; for, if I were to accept everything, then no truth would escape my acceptance. The trouble is, neither of these simple objectives of acceptance is the stuff of which knowledge is made.

What is needed is acceptance that aims at accepting something exactly in those cases in which it is true and not otherwise. To put the matter in a formula, the relevant aim of acceptance is to accept that *p* if and only if it is true that *p*.

Truth

In order to understand this condition, we need to have some account of truth. We have already noted the following condition of knowledge:

(iT) If *S* knows that *p*, then it is true that *p*.

Thus, it is appropriate to inquire into the nature of truth to understand knowledge. Unfortunately, the notion of truth is shrouded in controversy and paradox. For example, philosophers disagree about what sort of objects are true or false. One natural suggestion is that it is sentences, declarative sentences—'the structure of the human genome is a double helix,' for example—that are true or false.

Let us consider the attempt to provide a theory of when sentences are true or false. A complete theory or definition of truth for sentences would fill in the right-hand side of the schema,

X is true if and only if . . .

and thus tell us the conditions under which each sentence is true. This would give us an answer to the question, What is truth? In fact, no such general theory of truth is possible. Why?

Suppose we attempt to give a general theory, call it *T*, of the conditions under which sentences are true. Accordingly, *T* informs us, at least minimally, of the conditions under which sentences are true. But now a nasty

question arises. Under what conditions is *T* itself true? Either *T* tells us the conditions under which *T* is itself true or it does not. If it does not, then *T* is not a *general* theory of truth because it does not tell us the conditions under which one sentence, *T* itself, is true. So, to be complete, *T* must tell us about the conditions under which *T* itself is true. But then it must refer to itself. There are some sentences that refer to themselves and lead to a very famous paradox. These paradoxes lead some philosophers to consider self-reference problematic and to refuse to assign truth value to such sentences. We shall not follow the paradoxes to this radical conclusion, but let us consider the paradox to decide how to deal with it.

The paradox may be formulated in terms of a sentence that says of itself that it is false as follows:

S. Sentence *S* is false.

If sentence *S* is false, then since it says sentence *S* is false, what *S* says is true, and sentence *S* is true. Of course, if *S* is true, then since it says *S* is false, *S* must be false. *S* is, therefore, true if and only if it is false.

The paradox might seem to arise simply because *S* refers to itself, but the paradox runs deeper. Consider the following two sentences:

A. Sentence *B* is true

and

B. Sentence *A* is false.

Sentence *A* refers to *B*, not to itself, while *B* refers to *A*, not to itself, and yet sentence *A* is true if and only if it is false, as the reader can easily determine. The crux is that the paradox arises because sentences of a language can be used to speak about sentences of the same language and say things about their truth or falsity.

These paradoxes should not be thought of as simple logical puzzles. Every solution has problematic consequences. It is possible to lay down some rule that excludes all such sentences from the domain of legitimate discourse. One might, for example, formulate rules for legitimate or well formed sentences of a language, as Alfred Tarski did,[1] which would exclude the paradoxical sentences as not well formed. As long as we speak according to the rules, no paradox will arise, but the remark that sentences like *S*, *A*, and *B* are not well formed itself leads to paradox.

Consider the sentence

C. Sentence *C* is either false or not well formed.

Brief reflection is required to note that sentence *C* is true if and only if it is either false or not well formed. A follower of Tarski who insists that C is not well formed is then in the peculiar position of saying something about *C* which, if true, would appear to imply that *C* itself is true. For the follower says

Sentence *C* is not well formed

which seems to imply that

Sentence *C* is either false or not well formed

which, of course, is *C* itself.

Moreover, the new paradox arises for any way of characterizing the paradoxical sentences, for example, as meaningless or as indeterminate, that is, neither true nor false. We need only substitute for the expression "not well formed" in *C* our preferred characterization of the paradoxical sentence to obtain a new paradox. In fact, we can formulate a generic paradox elicitor by replacing "not well formed" with a variable "*Q*" to obtain

CV. Sentence *CV* is either false or *Q*

which will yield a paradox by substituting one's favored characterization of the paradoxical sentences for the variable *Q*.

One way to escape this paradox is to refuse to utter the paradoxical sentences or say anything about them. One may put one's hand over one's mouth in silence. Silence may be the better part of dialectical valor, but it provides meager enlightenment. Furthermore, the sentences that are paradoxical refer to themselves or enter into loops of reference that lead from the sentence back to itself through the reference of other sentences. But such loops of reference do not seem to be a basis for precluding the assignment of truth to sentences generally. Consider, for example, the following two sentences:

SR. SR refers to itself

and

UT. All true sentences are true.

Both of these sentences are sentences we should wish to say are true rather than be forced into silence concerning their truth value, but both refer to themselves. Sentence *SR* obviously refers to itself and *UT* refers to itself if it

is true, for it refers to all true sentences. Thus, we want to allow that there are true sentences that refer to themselves.

Moreover, the paradoxes have an important consequence. They show that we cannot accept a very simple and minimal theory about the truth of sentences. We might call it the *disquotational* theory of truth because it gives an account of truth by dropping quotation marks around a sentence, by disquoting a sentence. One instance of the theory is

> "Chisholm is a philosopher" is true if and only if Chisholm is a philosopher

and the general form consists of sentences resulting from substituting the same declarative sentence for "*X*" when it occurs between quotation marks and at the end of the formula in

> "*X*" is true if and only if *X*.

Unfortunately, the disquotational theory of truth, once put forth as a condition of adequacy for a definition of truth by Tarski,[2] must be given up in an unrestricted form applied to a natural language, such as English, which contains the paradoxical sentences, as Tarski himself noted. (It may be satisfied in artificial languages of the sort that Tarski constructed, which, unlike English, do not allow the formation of paradoxical sentences, but they also fail to allow the formation of sentences that refer to themselves in the way that *SR* and *UT* do.) Since our pretheoretical understanding of truth is probably based on the acceptance of something like the unrestricted general disquotational theory, it is important to notice that it leads to paradox and cannot be sustained.

Thus, a complete theory of truth is impossible. Tarski noted that the attempt to formulate a complete theory of truth for a language within the language itself would lead to paradox.[3] This is a technical result of major importance that contains a metaphysical insight of equal importance. It is that the attempt to give a complete account of the relationship between language and the world within language is doomed to failure. The paradoxes exhibit the failure in cases in which language is both the subject and vehicle of discourse. What should we conclude about truth? It is a notion that cannot be defined. We cannot give any perfectly general definition of truth that specifies the necessary and sufficient conditions for the truth of every true sentence.

Does that mean that truth is mysterious? Only if you think that anything that cannot be defined is mysterious. Some have thought that the word "red" cannot be defined because it is a simple notion whose meaning can only be understood from experience. Suppose the word "red" cannot be

defined. Would you conclude that redness is mysterious? Surely not. Notice, moreover, that for any sentence that does not refer to itself either directly or indirectly, the minimal theory of truth applies. For most of the sentences of a language, though not all, the minimal theory of truth specifies a condition of the truth of the sentence.[4] Moreover, we may allow the assignment of truth or falsity to other sentences such as *SR* and *UT* which we may consider true without paradox provided we give up the general condition of adequacy from Tarski.[5]

The paradoxes transfer from words to ideas, from sentences to acceptings and thus become directly relevant to our inquiry. We can obtain paradoxes from an analogy to the disquotational theory applied to acceptance. It is natural to assume that

> What *S* accepts, that Chisholm is a philosopher, is true if and only if *S* accepts that Chisholm is a philosopher and Chisholm is a philosopher

and in general

> (G) What *S* accepts, that *p*, is true if and only if *S* accepts that *p* and *p*,

but this leads to paradox. Consider the following:

> (F) What the only person in room 226 is now accepting is false.[6]

Suppose that, without being aware of the fact, I am that person, that is, the only person in room 226. Given (G) and my being the person in 226, my accepting that (F) is true if and only if I accept that (F) and what I am accepting is false. So, my accepting what I do is true if and only if it is false.

What is the upshot of this discussion for our inquiry? First of all, we must proceed without assuming that we can give any general theory or definition of truth pertaining to either sentences or acceptances we have considered. We may, nevertheless, retain the proposal that the sort of acceptance germane to knowledge is acceptance that aims at truth. The paradoxes concerning truth do not imply that it is paradoxical to aim at accepting *p* if and only if it is true that *p*. We may even formulate the objective without using the word "true" at all by saying that the aim of acceptance is to accept that *p* if and only if *p*. This objective, however formulated, is not rendered paradoxical by the truth paradoxes. The paradoxes simply are cases in which the objective cannot be attained.

Consider the claim (F) above. Given the objective of accepting that *p* if and only if *p*, I have the objective of accepting that what the only person in room 226 is now accepting is false if and only if what the only person in

room 226 is now accepting is false. When I am the only person in room 226 and accept that what the only person in 226 is now accepting is false, I shall be frustrated in my attempt to obtain my objective because what the only person in 226 is now accepting is true if and only if it is false. In general, paradoxical acceptances will yield cases in which there is no way in principle of fulfilling the objective in question and, therefore, in those cases the objective has no application.

Moreover, paradoxical acceptances will, as we have noted, arise from the attempt to accept a complete theory of when what we accept is true, of the relationship between acceptance and reality. This should not be seen as a counsel of despair, however. Most instances of (G) do not lead to paradox, indeed, almost none do, and the general assumption (G) is nonproblematic with respect to such instances. It supplies us with at least a minimal account of truth for those instances. Similar remarks apply to the absolute theory of truth

(AT) It is true that p if and only if p

considered above. The claim (F) above will generate paradox when substituted for the variable 'p' in (AT) as well as in (G). But most substitutions in (G) and (AT) will lead to nonparadoxical equivalences, indeed, to equivalences that are necessarily true.

Moreover, the account of truth offered for nonparadoxical instances allows for further theoretical articulation. For example, one might claim for such instances that S's acceptance that p is true if and only if S's acceptance that p corresponds to the fact that p. Given the correctness of the assumption (G), however, it must be the case that S's accepting that p corresponds to the fact that p if and only if S accepts that p and p. In that case, S's acceptance that p corresponds to the fact that p just in case S accepts that p and p. Thus, it appears that an account of what it is to accept that p, what it is for a mental state of acceptance to be an acceptance that p, to have that content rather than another, yields at least a minimal account of correspondence.

To return to our example, if we can obtain an account of what makes a state of accepting something a state of accepting that Chisholm is a philosopher, of what gives it that specific content, we shall thereby have obtained a minimal account of what it is for my accepting that Chisholm is a philosopher to correspond to the fact that Chisholm is a philosopher. When my state of accepting has the content that Chisholm is a philosopher, and Chisholm is a philosopher, my acceptance corresponds to the fact that Chisholm is a philosopher. Thus, we may conclude that a theory of the content of acceptance and thought generally yields a minimal theory of correspondence for nonproblematic cases.

Acceptance and Knowledge

Let us now turn to a reconsideration of the acceptance condition of knowledge contained in the following conditional:

(iA) If *S* knows that *p*, then *S* accepts that *p*.

As we noted in the first chapter, the more common proposal is the following conditional:

(iB) If *S* knows that *p*, then *S* believes that *p*.

Conditional (iB) has been a battleground of controversy, and, though we have replaced (iB) with (iA), it is reasonable to assume that some of the objections raised against the former would be called into battle against the latter as well, for we have said that acceptance of the required sort may be accompanied by a kind of belief so that

(A) If *S* accepts that *p*, then *S* believes that *p*

may be expected to hold true in most instances. It will, therefore, be useful to explore the objections to conditional (iB) at the outset to defend our proposal.

Two kinds of arguments have been employed in the effort to refute (iB). The first depends on certain facts of linguistic usage. For example, it makes sense and is sometimes quite correct to say 'I do not believe that; I know it' or 'she does not believe that; she knows it.' This kind of argument represents an attempt to show that it is inconsistent to say both that a person knows that *p* and that she believes that *p* and, consequently, that the former does not imply the latter.[7] The second form of argument is less ambitious and consists of offering a counterexample to (iB), the favored kind being one in which a person gives correct answers to questions she is asked without believing that her answers are correct.

The Consistency of Knowledge and Belief

Let us consider the first form of argument. From the fact that it makes sense and is even correct to say 'I do not believe that; I know it' or 'he does not believe that; he knows it,' it hardly follows that the thesis (iB) is false. The reason it makes sense to say these things is to be found in the study of rhetoric rather than logic. It makes sense to say, 'I do not believe that; I know it,' not because it is logically inconsistent to say that a person believes

what she knows but rather because this is an emphatic way of saying, 'I do not only believe that: I know it.' To say the latter, however, is quite consistent with conceding that the person referred to does believe, though not only believe, what she is said to know. Similar remarks apply to the locution, 'She does not believe that; she knows it.' An exact analogy to these cases is one in which it makes sense to say, 'That is not a house, it is a mansion,' and the reason it makes sense is not that it is logically inconsistent to say that a house is a mansion but rather that this is an emphatic way of saying, 'That is not only a house, it is a mansion.' Indeed, that something is a mansion entails that it is a house. Once the rhetoric of emphatic utterance is understood, the logic is left untouched.

It is worth noting how the replacement of acceptance for belief makes the objection less plausible. It is odd to say 'I do not accept that; I know it' or 'she does not accept that; she knows it,' though one would probably be understood. Such expressions are less natural, however, which suggests that the contrast between knowledge and acceptance is less salient than the contrast between knowledge and belief. Acceptance makes a more natural ingredient of knowledge than belief.

'I Know' As a Performative Utterance

A more sophisticated attack on (iB) resting upon considerations of linguistic usage is derived from the writings of J. L. Austin. In a famous passage Austin compares the locution 'I promise' to the locution 'I know.'[8] His basic contention is that uttering such words is the performance of a certain ritual that alters one's relations to others. As he puts it, "When I say, 'I promise,' I have not merely told you what I intend to do, but by using this formula (performing this ritual), I have bound myself to others, and staked my reputation in a new way."[9] Similarly, Austin says, "When I say 'I know *S* is *P*,' I give others my word: I give others my authority for saying that '*S* is *P*.'"[10] Austin goes on to remark, "To suppose that 'I know' is a descriptive phrase is only one example of the descriptive fallacy, so common in philosophy. Even if some language is now purely descriptive, language was not in origin so, and much of it is still not so. Utterance of obvious ritual phrases, in the appropriate circumstances, is not describing the action we are doing, but doing it."[11]

Even if Austin is correct in declaring that when I say 'I know that *S* is *P*,' I give others my authority for saying that *S* is *P*, the performance of this act is perfectly consistent with describing oneself as accepting with justification that *S* is *P*. Indeed, the assumption that I am making such a descriptive claim about myself when I say 'I know that *S* is *P*' helps to explain the way in which I give my authority for saying that *S* is *P* by saying 'I know that *S* is *P*.' If I were not claiming to be justified in accepting that *S* is *P* when I say 'I know that *S* is *P*,' then why in the world should my saying 'I know that *S*

is *P'* be taken as giving my authority for saying that *S* is *P?* It might be more reasonably be taken as an expression of opinionless agnosticism.

Knowing Implies Believing: An Alleged Counterexample

Let us now consider the second form of argument directed against (iB), the attempt to produce a counterexample. The best instance of such an argument that I have found is in an article by Colin Radford.[12] Radford's alleged counterexample to (iB) would, if correct, succeed as well against (iA). The distinction between belief and acceptance is not germane to the example. It concerns a man, John, who protests quite sincerely that he does not know any English history but when quizzed is able to answer some history questions correctly, for example, ones concerning the dates of the death of Elizabeth I and James I. John also makes some mistakes—indeed, he misses the mark more often than he hits it—and he cannot tell when he is right and when wrong. John thinks he is guessing all along. Because he thinks he is guessing, he is not inclined to believe or, we might add, to accept that his answers are correct.

Nevertheless, Radford contends we should say that John knows some history. For example, John gives the correct answer to the question concerning the year of Elizabeth's death, and so he knows the answer: Elizabeth died in 1603. Radford bolsters this contention by asking us to suppose that John has previously learned these dates and, consequently, that the reason he gives correct answers is that he remembers them.

Must we concede that John knows that Elizabeth died in 1603, even though he does not believe or accept that this is so? We may resist the inclination to do so, if any exists, by arguing that John does not know that Elizabeth died in 1603. One strategy would be simply to deny that John *knows* the correct answer, though, to be sure, he gives the correct answer. It is, however, natural enough to say that John knows the correct answer, and, consequently, it is useful to attack the argument by means of a counterargument.

The crucial premise of such a counterargument is that, though John knows some correct answers, he does not know that these answers are correct. This is shown by the fact that he has no idea which of his answers are correct. He considers himself to be untrustworthy in distinguishing the correct answers from the incorrect ones. He is not in a position to evaluate whether his state fulfills the objective of obtaining truth and avoiding error. Thus, though John answers the question concerning the death of Elizabeth correctly, he does not know that his answer is correct. But what does he need to know in order to know that his answer, that Elizabeth died in 1603, is correct? To know it is correct, all he needs to know is that Elizabeth did, in fact, die in 1603. If he knew that she died in 1603, then he would also know that his answer is correct, for he knows what he has answered; but he

does not know that his answer is correct. Therefore, John does not know that Elizabeth died in 1603.

We have, from one example, elicited contradictory conclusions. Let us look at the arguments side by side. Put schematically, Radford's argument is as follows:

1. John knows the correct answer to the question.
2. The correct answer to the question is that Elizabeth died in 1603.
3. If John knows the correct answer and the correct answer is that Elizabeth died in 1603, then John knows that Elizabeth died in 1603.

Therefore,

4. John knows that Elizabeth died in 1603.

The opposing argument is as follows:

1. John does not know that his answer is correct.
2. John's answer is that Elizabeth died in 1603.
3. If John does not know that his answer is correct and John's answer is that Elizabeth died in 1603, then John does not know that Elizabeth died in 1603.

Therefore,

4. John does not know that Elizabeth died in 1603.

The second argument is equally persuasive and, moreover, there is no equivocation in the word 'know' in the conclusions of these arguments to lessen the force of the contradiction. With such contradictory conclusions cogently defended, must we concede that the concept of knowledge is contradictory? Is knowledge impossible?

Fortunately, there is no need to concede the impossibility of knowledge. Instead, we may reject premise (3) of Radford's argument. We may say that John knows the correct answer—that Elizabeth died in 1603—but deny that John knows that Elizabeth died in 1603. To see why, consider that we might well say that a woman knows the correct answer to a question about the date of Elizabeth's death even though she is guessing and thus does not know that Elizabeth died in 1603. Imagine a woman, Alice, who is on a quiz program and is asked the date of Elizabeth's death. She answers '1603.' You, not having heard the answer, ask me, 'Did Alice know the correct answer?' To this question I could reply in the affirmative. Moreover, it might not matter whether I thought Alice was guessing or not. When the

quiz master says, 'Alice, if you know the answer to the question I am about to ask, you will win that red Ford,' he does not intend to withhold the Ford if Alice guesses correctly. On the contrary, in this context to give the correct answer is to know the correct answer. When you asked, 'Did Alice know the correct answer?' I could have answered, 'Yes, but I think it was just a lucky guess.' There is no question of whether Alice knows the correct answer once she gives it. She knew. Away she drives.

Thus Alice, like John, knows the correct answer, that Elizabeth died in 1603. Alice knew just as John did, even though Alice was guessing. Though we concede that Alice knew the correct answer, we should want to insist that she did not know that Elizabeth died in 1603. A lucky guess that *p* is not a case of knowing that *p*. The preceding argument shows that there are contexts in which it would be acceptable to say that a person knows the correct answer, which is that *p*, but would be clearly false to say the person knows that *p*.

An Objection: Remembrance Without Belief

Radford has objected to the preceding argument on the grounds that the question has been begged against him.[13] In support of this, he appeals to the consideration that one may remember that *p*, and hence know that *p* from memory, when one does not know that one knows that *p*, believe that one knows that *p*, or even believe that *p*. Indeed, the example concerning John, as spelled out in detail, is one in which he has previously learned that Elizabeth died in 1603, though he has forgotten having learned it, and thus gives the right answer because he remembers what he once learned. He remembers even though he does not know or believe himself to have done so, and consequently believes he is guessing. Since John is remembering, he knows that *p*, even though he does not believe he knows, thinks he is guessing, and does not believe that *p*.

Borderline Cases of Knowledge

What are we to say to these arguments against (iB) and (iA)? The most direct reply is one conceded by Radford. At one point, Radford admits that his examples are borderline cases of knowledge.[14] This is precisely the case, but what is a borderline case? To say that a case is borderline means there are considerations in favor of applying the term, and equally strong considerations in favor of not applying it. For example, if we see something that is very similar in color to many red things, so much so that this is a reason for saying that the object is red; but at the same time it is very similar to things that are orange and not red, so much so that this is a reason for saying that the thing is not red, then we have a borderline case of something red. Such cases abound.

For most terms of everyday speech, we can expect to find that the term applies without doubt or controversy in a large number of cases, and that it also clearly fails to apply in many cases. On the other hand, in between these cases there are examples of things where it is unclear whether or not the term applies, no matter how much we know about the example. Here we are very likely to conclude that the case is borderline. Debate on whether the term applies in such cases can produce perplexing arguments and confusing speculation, but no one can win because the case is precisely one in which the application of the term is not fixed. As Stephan Körner has suggested, such terms are inherently inexact and, therefore, the decision to apply or not to apply the term in borderline cases is a matter of choice.[15]

If we wish to defend (iB) and (iA) by maintaining that the case of John is borderline for the application of the word 'know,' then two tasks remain. First, we must show that the case and others like it are genuinely borderline and, second, we must justify our choice of applying epistemic terms in the manner required for the truth of (iA).

Some argument for the conclusion that John's case is a borderline case of knowledge has already been given in the earlier presentation of two persuasive arguments, one yielding the conclusion that John does not know that Elizabeth died in 1603, and the other the exact opposite. Such arguments and counterarguments concerning these examples show them to be borderline cases. It is precisely like arguing about whether something is red when it is as close to red as it is to orange; and no argument can settle that.

Knowledge Implies Acceptance

Given, as we have argued, that the case of John is borderline, how can we justify refusing to apply the term 'know' in such cases? The appropriate justification is theoretical, one concerning the role of acceptance and the evaluation of information in knowledge and justification. A person may be said to possess the information that Elizabeth died in 1603 when he has retained it in memory, even when he cannot access it, but he does not, at a given time, know that the information is correct because he does not accept that it is and cannot use the information in critical reasoning. John possesses the information that Elizabeth died in 1603 in that it is retained in his memory. Since the information is retained in his memory, we say, when he produces an answer, that he remembered that Elizabeth died in 1603, even though he does not know that Elizabeth died in 1603. This is a case, of which there are many, where remembering that p does not logically imply knowing that p. The reason John does not know is that, though he possesses the information that p in memory, John cannot positively evaluate the truth of the information and does not know that the information is correct.

Some philosophers, Fred Dretske, for example, have assumed that if a person receives information that *p*, then one knows that *p*.[16] Others, like Radford, have assumed that if a person receives and retains the information that *p*, then the person knows that *p*. Both views have the same defect. There is an important distinction between receiving or retaining information and knowing that the received or retained information is correct. Only when one can evaluate the truth of the information and, consequently, know that the information one receives or retains is correct does one have knowledge. Memory gives us knowledge of the information retained when we can positively evaluate the correctness of the information. Clear and distinct memory carries positive evaluation of the correctness of the information remembered with it and provides us with knowledge of it. When the capacity to evaluate the retained information is lost, the knowledge vanishes with the loss of capacity to evaluate whether the retained information is correct.

Imagine, for example, that Mary had been told that Elizabeth died in 1603 by Peter, who is notoriously untrustworthy in such matters. Imagine further that Mary retained that report in memory only because she had accepted what Peter had told her, in spite of knowing him to be untrustworthy in such matters. It would be obvious that Mary did not know that Elizabeth died in 1603 when Peter told her because she did not know that the report she received was correct.

It should be equally obvious that John does not know that Elizabeth died in 1603, when that information is retained in memory, because he does not know that the information is correct. John, unlike Mary, might have once known that Elizabeth died in 1603 because he once knew that the information was correct. Now, though the information is retained in memory, John no longer knows that the information is correct because he cannot evaluate whether the information is correct or not. That is why John does not know that Elizabeth died in 1603. John does not know this because he does not know that the information he possesses is correct. He does not know that the information he possesses is correct because he cannot positively evaluate that it is correct and, consequently, does not accept it.

These reflections provide us with the basis for a perfectly general proof that knowledge implies acceptance. The proof is as follows:

1. If a person does not accept that *p*, then the person does not accept the information that *p*.
2. If a person does not accept the information that *p*, then the person does not know that the information that *p* is correct.
3. If a person does not know that the information that *p* is correct, then the person does not know that *p*.

Therefore,

4. If a person does not accept that *p*, then the person does not know that *p*.

This is equivalent to the acceptance condition,

(iA) If *S* knows that *p*, then *S* accepts that *p*

with which we began.

The premises of the argument, once made explicit, may seem so obvious as to require no proof. They do, however, reflect our concern with a kind of acceptance and knowledge consisting of the recognition of information and positive evaluation of it. Acceptance of information is not a sufficient condition for knowing that the information received is correct, but it is necessary. Consequently, the failure to accept information results in failure to know that the information is correct. The problem is clear in the case of John and many similar cases in which information originally known to be correct is retained but the knowledge of the correctness of it is lost. Some information that *p* is retained in memory, but, at some point in time, the capacity to positively evaluate the information that *p* is lost and with it the acceptance of *p* and knowledge that *p*. In short, it does not suffice for knowledge that some information is stored in memory as a result of communication or perception. One must, in addition, have the appropriate sort of evaluation of the information and acceptance of it.

Knowledge and the Functional Role of Acceptance

The foregoing remarks may be construed as further elucidation of the kind of knowledge that is the object of our study. There are living beings, such as a gullible child, as well as machines, such as an answering device, that receive and retain information. If we know enough about those beings and machines, they may be a source of information and knowledge for us. They, however, do not know that the information they possess, retain, and transmit is correct because they cannot evaluate it. We, on the contrary, not only receive and transmit information, we evaluate it. We accept some but not all of what we receive and know that some but not all of the information is correct.

The nature and role of acceptance in knowledge requires some clarification. Acceptance is the sort of mental state that has a specific sort of role, a functional role, in thought, inference, and action. When a person accepts

that *p,* he or she will reason in a certain manner and perform certain actions assuming the truth of *p*. Thus, if a person accepts that *p,* then the person will be ready to affirm that *p* or to concede that *p* in the appropriate circumstances and to use *p* to justify other conclusions. They will also be ready to justify the claim that *p*. If they accept information received from the senses or retained in memory, they will regard such information as correct and proceed accordingly in thought and action. The reluctance of John to affirm that Elizabeth died in 1603 reveals ignorance and lack of acceptance of the information. Acceptance of *p* sometimes arises from considered judgment that *p,* but a functionally similar state of judgment may arise in other ways. To accept that *p* involves a positive evaluation of the information that *p*. To accept the information that *p* implies a readiness in the appropriate circumstances to think, reason, and act on the assumption that the information is correct.

Acceptance, though it may often accompany positively evaluated belief, may be contrasted with simple belief. There is an illuminating analogy between simple belief contrasted with acceptance, which incorporate positive evaluation of a belief, on the one hand, and simple desire contrasted with what we might call *preference*, which involves positive evaluation of the desire, on the other hand. Desires, like beliefs, arise in us naturally without our bidding and often against our will. As we consider our desires, we find that satisfying some of them is contrary to our purposes, while the satisfaction of others accords with our purposes. Those that accord with our purposes are ones we evaluate positively and prefer to satisfy, while those that conflict with our purposes we refuse to certify and do not prefer to satisfy. When we act and reflect rationally in terms of our purposes, we are directed by those desires we prefer to satisfy and not by the others. We may, of course, be overwhelmed by desires that are contrary to our preferences, but preference is the guide of reason.

Similarly, some of our beliefs accord with our purposes of obtaining truth and avoiding error, and those we positively evaluate and accept for these purposes, whereas other beliefs conflict with these purposes and do not receive the certification of acceptance. We should not, however, expect that desires we do not prefer to satisfy or beliefs we do not accept automatically vanish. The mind of evaluation is, I have suggested,[17] a higher-level mind, a metamind of reason, of preference and acceptance, contrasted with a mind of belief and desire which, though it may be influenced, is an independent, lower-level mind of its own. It is acceptance that is requisite to knowledge because knowledge supplies the premises of critical reasoning requiring positive evaluation in terms of the purposes of obtaining truth and avoiding errors. These are the purposes of acceptance.

The result of our argument is that we shall resolve the borderline cases considered by saying the subjects lack knowledge. For some purposes of everyday speech, we can afford the semantic imperfection that yields the sort of contradictory conclusions we have derived from the study of John, but a satisfactory theory must eliminate such imperfection to avoid contradiction. We shall require that the epistemic terms in question carry the implication of acceptance, for this will enable us to extricate our employment of the terms from the contradictions noted earlier. Such a requirement leads directly to the conclusion that John does not know that *p* when he says he does not know because he lacks acceptance that *p*. By so doing, we are not dogmatically ruling out the possibility that some theory of knowledge might be constructed which would rule in the opposite direction. We should welcome the development of such a theory. However, as we affirmed earlier, our concern is to present a theory of knowledge that plays the appropriate role in critical reasoning—in justification, confirmation, and refutation. Such a theory of knowledge and justification seeks to explain how we are justified in accepting information as correct and claiming to know that it is. Examples of alleged knowledge in which a person does not know that the information he accepts is correct may be of some philosophical interest but such knowledge falls outside the concern of knowledge used in a way that is characteristically human in critical reasoning and the life of reason.

In defense of the foregoing restriction, it should be added that our decision to require the implication of acceptance of the information that *p*, and accompanying readiness to affirm that *p* in the appropriate circumstances when a person knows that *p*, is nothing arbitrary or idiosyncratic. It is warranted by the fact that our edifice of scientific knowledge and practical wisdom depends upon the social context in which criticism and defense determine which claims are to be employed as the postulates of scientific systems and the information for practical decisions. In such contexts, a person who admits ignorance is taken at her word; for such a person is not willing to make the sort of epistemic commitments that would enable us to check her cognitive credentials. Of course, we may well be interested in her reasons, if she have any, for conjecturing what she does, but this is quite different from asking whether she is justified in accepting something as correct information and claiming to have knowledge. An affirmative answer to that question not only shows that the person has knowledge, it also transfers that knowledge to those who understand the justification and apprehend its merits. Our theoretical concern with critical reasoning, as well as our attempt to explain how such reasoning succeeds, warrants our decision to rule that knowledge must involve the forms of acceptance cited above.

Memory Without Knowledge

At this point, we must honestly face the question, Why do such cases seem to be examples of knowledge to philosophers of merit? The answer is that we often ascribe knowledge to others in order to explain, in a commonsense manner, why they are correct. In the case of John, for example, someone is in a position to say something that warrants our concluding that the person was in a position to give the right answer. It is not just a matter of luck that he can give a correct answer. On the contrary, it is because he acquired knowledge in the past and retains information in memory that it is possible for him to come up with the correct answer now. If we suppose that the knowledge has vanished, then it seems difficult to explain how he can now be in a position to give the correct answer. For the purposes of explaining how he can now be in a position to give a correct answer, we conclude that the knowledge has not vanished. He still knows.

This very natural way of explaining correct answers is, however, highly defective. Something extremely important has been lost by the person in spite of such answers, namely, the positive evaluation and acceptance of the information. In the absence of such acceptance, the person does not know the information retained is correct. The resulting refusal to defend the claim to know abrogates the use of the claim in critical reasoning. The difference between the person who accepts the information that *p* and is ready to defend the claim that she knows, on the one hand, and a person who does not accept the information that *p* and thinks she is just guessing that *p*, on the other, is sufficiently great measured in terms of usefulness for critical reasoning so that we may justifiably mark the distinction by refusing to say that the latter knows that *p*. By doing so, we shall in no way prevent ourselves from explaining how a person can manage to give the correct answer, that *p*, without knowing that *p*.

The latter contention is supported by the application of another argument derived from David Armstrong.[18] He asks us to imagine a case in which a man, asked about the date of the death of Elizabeth, answers, 'Elizabeth died in 1306.' Now this answer, though incorrect, is sufficiently similar to the correct answer, 1603, so that we can see that it is not a mere matter of chance that he gave this answer. We cannot explain this man's answer by affirming that he knows that Elizabeth died in 1603, for, since the numbers six and three having been transposed in his memory, he does not know that Elizabeth died in 1603. (To clarify this we may even imagine that when asked, 'Are you sure it was not 1603?' the man replies that he is sure it is 1306.) Just as it would be incorrect to explain why this man gives the answer he does by affirming that he knows the date of Elizabeth's death, so explaining why John and the others are able to give a correct answer by affirming that they know, would be otiose. A theory of memory that explains

how memory produces the results it does, whether correct or incorrect, does not require us to assume that one knows one's answer is correct whenever memory enables one to produce a correct answer. To do so would be to lump together the case in which memory only functions well enough to yield a correct answer with those cases in which it produces acceptance and justification.

Our contention is that sometimes memory is good enough to give us a correct answer when it is not good enough to give us knowledge that the answer is correct. When memory does not function well enough for us to know that the information we have retained is correct, then we are not in a position either to claim to know or to justify a claim to know. Thus, we wish to distinguish sharply between those cases in which memory serves us so well that we know the information retained in memory is correct from those in which the information retained in memory is not accepted as correct and, consequently, is not known to be correct. That there is a distinction cannot be doubted, and, for the purpose of constructing a theory to explain epistemic justification and account for critical reasoning, we need only count as knowledge those cases in which a person accepts the information in question and knows that the information is correct. Such cases have top value in the epistemic marketplace of the life reason, and all the rest may be discounted without explanatory loss.

Introduction to the Literature

The best anthology of articles on truth is *Recent Essays on Truth and the Liar Paradox,* by Robert Martin. The two best studies of the subject of truth are *Truth,* by Paul Horwich; and *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth,* by Vann McGee. The classic article on truth is by Alfred Tarski, "The Semantic Conception of Truth."

The articles with important arguments to show that knowing does not entail believing include "Knowledge—By Examples," by Colin Radford; and "Other Minds," by J. L. Austin. For an important and clear discussion of the topic, see David Armstrong, "Does Knowledge Entail Belief?" The claim that knowledge consists of the receiving of information is articulated in detail by Fred Dretske in *Knowledge and the Flow of Information.* A fuller account of the distinction between acceptance and belief is to be found in *Self-Trust: A Study of Reason, Knowledge and Autonomy,* by Keith Lehrer.

Notes

1. Alfred Tarski, "The Semantic Conception of Truth," in *Semantics and the Philosophy of Language,* ed. L. Linsky (Urbana: University of Illinois Press, 1952);

Tarski, "The Concept of Truth in Formalized Languages," in *Logic, Semantics, and Metamathematics* (Oxford: Clarendon, 1969).

2. Tarski, "Semantic Conception of Truth."

3. Ibid.

4. Cf. Paul Horwich, *Truth* (Oxford: Blackwell, 1990).

5. Vann McGee, *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth* (New York: Hackett, 1990).

6. This example is due to Marian David.

7. This kind of argument is more popular with philosophers than the literature of the field reveals, and it would be hard to say who first formulated it. The late Austin Duncan-Jones remarked, "Yet obviously, if 'I think *p*' is used in the colloquial way, as equivalent to, 'I am inclined to think *p*,' it doesn't follow from 'I think *p*' at all but is consistent with it." In a footnote on the same page he remarks in support that we say, 'I don't think, I know.' These remarks appear in "Further Questions About 'Know' and 'Think,'" in *Philosophy and Analysis*, ed. Margaret Macdonald (Oxford: Blackwell, 1966), 97.

8. J. L. Austin, "Other Minds," in *Logic and Language*, ed. A. G. N. Flew, 2d series (Oxford: Blackwell, 1959), 123–58.

9. Ibid., 144.

10. Ibid.

11. Ibid., 146–47.

12. Colin Radford, "Knowledge—By Examples," *Analysis* 27 (1966): 1–11. Much of Radford's attack in this article is directed against the thesis that a person knows that *p* only if he is sure or feels sure that *p*.

13. Radford, "Analyzing 'Knows That,'" *Philosophical Quarterly* 20 (1970): 228–29.

14. Radford, "Knowledge—By Examples," 4 n. 1.

15. Stephan Körner, *Experience and Theory* (London: Routledge & Kegan Paul, 1966), 19–47.

16. Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981).

17. Keith Lehrer, *Self-Trust: A Study of Reason, Knowledge, and Autonomy* (Oxford: Clarendon, 1997), 1–24.

18. David Armstrong, "Does Knowledge Entail Belief?" *Proceedings of the Aristotelian Society* 70 (1969–70): 21–36.

3

The Foundation Theory: Infallible Foundationalism

KNOWLEDGE IMPLIES justified acceptance. We expressed this in the condition

(iJ) If *S* knows that *p*, then *S* is justified in accepting that *p*.

What makes us justified in accepting one thing rather than another? One answer, supplied by *foundation theory,* is that some beliefs, basic beliefs, are justified in themselves and constitute the foundation for the justification of everything else. In this chapter we shall examine the foundation theory of justification according to which all justification is based on self-justified or basic beliefs. It is traditional to formulate and discuss the foundation theory in terms of belief rather than acceptance, and we shall do so as well; but the kind of state required for knowledge is acceptance directed at obtaining truth and avoiding error in what is accepted. We succeed in this objective when we accept something if and only if what we accept is true.

Thus, for the foundation theory to succeed, the self-justified or basic beliefs must be things we are justified in accepting in our quest for truth and do not depend for their justification on anything else that we accept. Everything else that we are justified in accepting must ultimately be based on these basic beliefs, which provide us with the foundation for the edifice of justification and knowledge. Is such a theory tenable?

We must first notice that any correct theory of justification must share at least one fundamentally important tenet of the foundation theory, to wit,

that there are some things which we are justified in accepting without having proven or even argued that they are true. Clearly, whatever our capacity for argument, we have not, in fact, argued for the truth of everything we are justified in accepting. For example, no one but a few philosophers has ever argued that they have a headache or that they are thinking, but many who have never argued for such things have, nevertheless, been justified in accepting them. We must, therefore, agree with the foundationalist that we are justified in accepting some things without argument.

Infallible Versus Fallible Foundationalism

Disagreement arises when we consider answers to the question, How are we justified in accepting things without any argument to show that they are true? Acceptance aims at truth. If we accept something without any argument for the truth of it, how can we be justified in accepting it for the purpose of accepting what is true? It is to this question that the traditional foundationalist has provided an important answer, namely, that some beliefs *guarantee* their own truth. If my accepting something guarantees the truth of what I accept, then I am justified in accepting it for the purpose of obtaining truth and avoiding error. We are guaranteed success in our quest for truth and cannot fail. We might, therefore, call this *infallible* foundationalism. Assuming that there are beliefs that guarantee their own truth and that these suffice to justify us in accepting all that we are justified in accepting, infallible foundationalism provides a brilliant solution to the problem of explaining how we can obtain knowledge.

We shall soon turn to a detailed examination of the merits of infallible foundationalism, but we should first note that there have been foundationalists of other sorts. For reasons we shall soon consider, a foundationalist might despair of finding a sufficient quantity of beliefs that guarantee their own truth and settle for a more modest foundation of self-justified or basic beliefs that provide a *reason* for their acceptance but without a guarantee of their truth. Since such a reason is a fallible guide rather than a guarantee of truth, we might call a theory of this kind *fallible* foundationalism. It is characteristic of fallible foundationalism to allow that the reason a self-justified belief provides for acceptance may be overridden or defeated by other considerations and, therefore, that the reason for acceptance is a prima facie reason for acceptance.[1]

Fallible Foundations

Some recent philosophers have, indeed, claimed that basic beliefs are reasonable or evident in themselves without going on to claim that such beliefs guarantee their truth or are immune from refutation.[2] On such a founda-

tion theory, the basic beliefs are justified unless there is evidence to the contrary. In other words, they are prima facie justified, innocent in the court of justification, unless their justification is overridden. Some foundation theorists, who have denied that the justification of basic beliefs need provide any guarantee of truth, have gone so far as to deny that such justification is connected with truth in any way at all.[3] Such theories, though philosophically important, leave us with a dilemma. Either such justification is irrelevant to the truth of basic beliefs or it is relevant. If the justification is irrelevant to the truth of basic beliefs, then it is not the sort of justification needed to justify acceptance or to yield knowledge. The acceptance required for knowledge is acceptance that aims at truth. Therefore, no justification that is irrelevant to truth is adequate to justify such acceptance.

Suppose, then, that the foundation theorist maintains that the justification of basic beliefs is relevant to the truth of those beliefs, though it does not guarantee their truth. The foundation theorist must then explain how the justification is relevant to the truth of basic beliefs if it fails to guarantee their truth. If there is some risk of error, moreover, then the justification such basic beliefs possess must offset the risk, that is, it must make the risk worthwhile. The risk or probability of error infecting our basic beliefs must not be too high, or else we would not be justified in accepting those beliefs as our foundation. If, however, there is some risk of error in accepting a basic belief, how can we be justified in accepting the belief without any explanation of why the risk of error is acceptable?

The foundation theorist may simply postulate that we are justified in accepting certain basic beliefs and give no rational explanation for this claim. But, in fact, there is an explanation. We agree on intuitive grounds that we are justified in accepting the beliefs in question, and why do we think that the beliefs in question are justified? The explanation is obvious. It is because we believe that they are sufficiently unlikely to turn out to be false or, what is the same thing, sufficiently likely to be true.

We agree that we are justified in accepting the beliefs because of the probability of their truth, but why do we think the beliefs in question are so likely to be true? When one considers the candidates for such beliefs—introspective beliefs concerning one's present thoughts and sensations or cautious perceptual beliefs about simple qualities we see directly before us—the answer is apparent. We think that our powers of introspection and perception are very unlikely to lead us into error in such simple matters. That is the explanation of the justification of these basic beliefs.

So the justification for accepting these beliefs, if they fail to guarantee their own truth, implicitly depends on a theory we have concerning the reliability of our cognitive powers. This means, however, that the allegedly basic beliefs in question are justified by relation to other beliefs and are not genuinely basic. Such a theory is not a pure foundation theory. The al-

legedly basic beliefs must stand in the appropriate relation to other beliefs for their justification. Though the defender of a fallible foundation theory deserves a fuller hearing, it appears that only those foundation theories holding that basic beliefs guarantee their own truth are pure foundation theories. Fallible foundation theories inevitably appear tainted with a component of another theory, the coherence theory, perhaps, needed to defend the probability of basic beliefs.[4] We shall, therefore, turn to an examination of the pure version of the foundation theory, the infallible foundation theory, which provides us with justified beliefs whose truth is guaranteed.

The Foundation Theory in General

The infallible foundation theory, like any foundation theory, requires basic beliefs. What conditions must be met for a belief to be basic? The first is that a basic belief must be self-justified rather than justified by relation to other beliefs. The second is that the justification of all justified beliefs depends on the self-justification of basic beliefs. A theory of justification with these features is one in which there are basic beliefs that are self-justified and that justify all nonbasic beliefs.

Traditionally, the doctrine of empiricism has been associated with the foundation theory. According to empiricist theories of knowledge and justification, there are some empirical statements (for example, that I see something moving or, more cautiously, that it appears to me as though I see something moving) which constitute the content of basic beliefs. The belief that such statements are true is a self-justified belief. All beliefs that are justified are so because of the justification provided by accepting the empirical statements in question. Thus, the acceptance of such empirical statements is basic. Exactly how the empirical statements are construed depends on the empiricist in question. However, the empirical statements that constitute the content of basic beliefs have always been statements to the effect that some item in sense experience has or lacks some quality or relationship discernible by means of the senses. Thus, the empirical statements are statements of observation.

Empiricists have disagreed about the objects of sense experience. The item sensed may be conceived of as a physical thing, like a chair or a meter, or it may be construed as some more subjective entity, like an appearance or a sense datum. Moreover, they have disagreed about what makes such statements self-justified and about how basic beliefs justify other statements. They do agree that there are observation statements constituting the content of basic beliefs whose acceptance justifies all that is justified and, moreover, refutes all that is refuted.

Though empiricist epistemology is most commonly associated with a foundation theory, there is no logical restriction or, for that matter, histori-

cal limitation of foundation theories to empiricism. Rationalistic philosophies of knowledge, for example, that of Descartes, have been foundation theories. Such a rationalist maintains that a belief may be certified by reason as having characteristics that make it basic—indubitability, for example. A strict rationalism would hold that basic beliefs, and the justification they provide for other beliefs, are certified by reason alone.[5] Similarly, a strict empiricism would hold that basic beliefs and the justification they provide for other beliefs are certified by experience alone.

Few philosophers would contend that all justification is derived solely from reason or solely from experience. That a conclusion follows from premises is ascertained by reason, and what the objects of sense experience are like is ascertained by experience. Of course, reason may play a role in the latter, and experience in the former, but it would generally be conceded that if all people were deprived of reason, then no one would be justified in believing any conclusion to be a logical consequence of a premise. Similarly, if we were all deprived of our senses, then no one would be justified in believing there to be any objects of sense experience. These are obvious truths, mentioned only to illustrate how misguided it is to conceive of epistemology as the battleground between rationalism and empiricism.

The Foundation As a Guarantee of Truth

We are adopting an entirely different approach that cuts across traditional lines. Rationalists and empiricists often share a common conception that leads to a foundation theory. They conceive of justification as being a guarantee of truth. Empiricists think that experience can guarantee the truth of basic beliefs and rationalists think that reason is the guarantee of truth. Basic beliefs are basic because they cannot be false; their truth is guaranteed. With this initial guarantee of truth in basic beliefs, the next problem is how to extend this guarantee to other beliefs.

Our earlier analysis of knowledge offers a simple explanation of why this doctrine should be held. Since one condition of knowledge is truth, it follows that no belief constitutes knowledge unless it is true. Thus, if our justification fails to guarantee the truth of what we accept, then it may leave us with a false belief. In that case, we lack knowledge, so justification sufficient to ensure us knowledge must, some foundation theorists have argued, guarantee the truth of what we accept.[6]

Another motive for the doctrine of infallible foundationalism is a consequence of our account of acceptance. If the goal of acceptance is to accept something just in case it is true, then acceptance, which guarantees its own truth, provides us with a prophylactic against accepting something false. Thus, though a fallible foundation theorist might deny that we need a guarantee for the truth of basic beliefs,[7] a central thesis of the

traditional and infallible foundation theory was that basic beliefs are immune from error and refutation. If basic beliefs were erroneous and refutable, then all that was justified by basic beliefs, all that was built upon them in the edifice of justification, might be undone by error. The very foundation of all justification might prove unsound. If there is nothing to ensure that such basic beliefs are true, then, ipso facto, there is nothing to ensure the truth of those beliefs they justify. All justification might rest on a false foundation.

There is, consequently, another advantage of infallible foundationalism that might motivate a modern author, namely, that it provides a solution to the problem of justification depending on error formulated in the first chapter raised by Edmund Gettier. Recall that the initial formulation of the problem was one in which a person was justified in believing something false, that Mr. Nogot (who is in the class) owns a Ferrari, and deduces the true conclusion, which is, therefore, justified, that at least one person in the class owns a Ferrari. If infallible foundationalism is correct, there is the prospect of starting with justified beliefs that *must be true* because they are basic. If the justification of nonbasic beliefs results from *valid deduction* from such basic beliefs or any other relationship that guarantees their truth on the basis of the basic beliefs, then the truth of the nonbasic beliefs will be guaranteed by the basic beliefs. Such justification of basic or nonbasic beliefs will never yield a justified false belief. This will ensure that no justification essentially depends on error or is defeated by it. The primary advantage of an infallible foundation theory is that it incorporates a guarantee of truth into justification.

Incorrigible Foundations

A foundation theory alleging that basic beliefs guarantee their truth faces two problems. The first is to show that there are some basic beliefs which can guarantee their own truth. The second is to show how basic beliefs can guarantee the truth of other beliefs. Let us consider the first problem. Philosophers have maintained that some beliefs guarantee their own truth and are thus self-justified because they are *incorrigible.* We shall now examine the tenability of this thesis. What is meant by saying that a belief is incorrigible?

Let us begin with the intuitive notion that an incorrigible belief is one such that the person who has it cannot be mistaken in believing what she does. We are immediately faced with the tricky little word 'can,' a semantic chameleon. What are we to understand it to mean? We may wisely begin with a technical notion and then, should that prove insufficiently subtle, turn to some modification. Let us begin with the notion of *logical possibility.*

It is logically impossible that someone has a female brother or that some number is larger than itself. Logical impossibility is a familiar notion, though it is clear that some modal notion, possibility, for example, must be taken as primitive. So, if we assume the notion of a possible world, for example, then we can say that something is logically possible if and only if it obtains in a possible world, or, equivalently, something is logically impossible if and only if it does not obtain in any possible world. We sometimes speak of the logical impossibility of certain sentences. For example, we might say that the following sentence, 'John has a female brother,' is logically impossible. However, when we say such things we are speaking elliptically. It is what is *stated* by the sentence that is logically impossible. What the sentence 'John has a female brother' states is that John has a female brother, and that is logically impossible. This explains why the sentence is not true in any possible world and, therefore, is contradictory or, as philosophers say, analytically false because 'brother' may be analyzed as meaning 'male sibling.' We thus come full tilt to the controversial notion of *analyticity.*

It is notoriously difficult to provide any satisfactory definition or criterion of analyticity or related notions. Some philosophers thus disregard the notion of analyticity as a philosophical relic of semantic battles lost long ago. This conclusion is premature. Some logical notion, whether that of contradictoriness, possibility, or impossibility, may be taken as basic and undefined. Once we concede that logical possibility or some other logical notion must be taken as basic and undefined, we must also admit there are going to be cases in which it is difficult to ascertain whether something is logically impossible. This is partly because the distinction between logic and other areas of inquiry is not clearly drawn. Nevertheless, there are many cases in which the application of the concept will be sufficiently precise for useful employment, for example, to take a case we shall consider, in the claim that it is logically impossible that somebody should believe that he exists and not believe anything.

A Definition of Incorrigibility

Let us now define incorrigibility in terms of the concept of logical impossibility. We can say, roughly, that a belief is incorrigible if and only if it is logically impossible for the belief to be mistaken. More formally, the definition is as follows:

> The belief that p is incorrigible for S if and only if it is logically impossible that S believes that p and p is false.

Given this definition of incorrigibility, it follows immediately that if a person believes something and her belief is incorrigible, then what she be-

lieves is true. If it is impossible that she should believe that *p* and *p* should be false, then, given that she does believe that *p,* it follows that *p* is true. Hence, in this sense such beliefs guarantee their own truth.

A problem arises when we consider whether incorrigible beliefs as so defined are self-justified or, indeed, justified at all, for it is logically impossible that any person should be mistaken in believing anything which is logically necessary. By saying that something is logically necessary, we mean no more or less than that it is the denial of something logically impossible. Thus, it is logically impossible that two plus seven does not equal nine and, hence, logically necessary that two plus seven equals nine. This means, however, that it is logically impossible that a person should believe that two plus seven equals nine and be mistaken in her belief. The reason is that it is logically impossible that two plus seven should not equal nine.

A Counterexample

No matter how complicated or esoteric the arithmetical belief might be, it remains the case that if what is believed is logically necessary, then it is logically impossible that the belief should be false. Hence, the belief is incorrigible. For example, if a person believes that there is a one-to-one correspondence between the set of whole or natural numbers and the set of even numbers, then she believes something that is logically necessary, and her belief is incorrigible. If, however, she believes this for some foolish reason, for example, she believes that after a certain point in the series of numbers there are no more odd ones, then we would conclude that, on the basis of her reasoning, she surely could have been mistaken and was quite unjustified in her belief. It was pretty much a matter of luck that she was correct in her belief, and she certainly did not know that what she believed was true.

In this example the person could have been mistaken in some sense, even though it was logically impossible that she should have been. Thus, there appears to be some important sense of the expression 'could not have been mistaken' that our current specification of incorrigibility fails to capture. Moreover, this demonstrates that the justification we have for believing certain necessary truths in arithmetic, mathematics, and logic is not a simple consequence of the necessity of what is believed. A person may believe something that is a necessary truth without in any way knowing that her belief is true or even being justified in her belief. This argument proves that the logical impossibility of being mistaken does not suffice for justification when what is believed is an arithmetical, a mathematical, or a logical truth. When it is logically necessary that *p,* as it is in such cases, the logical impossibility of the conjunction that *S* believes that *p and* it is false that *p* is a direct consequence of the logical impossibility of the falsity of *p*. That *S* believes that *p* has nothing whatever to do with the incorrigibility of her belief in these cases.[8]

An Amended Definition of Incorrigibility

The solution to this problem is a simple amendment of the definition of incorrigibility, one that will, moreover, ensure that incorrigible beliefs achieve the goal of acceptance. We said that the objective of acceptance is to accept that *p* if and only if it is true that *p*. The definition of incorrigibility given above logically ensures that whenever a person's acceptance of *p* is incorrigible, then, if a person accepts that *p*, it is true that *p*, but this is only one objective of acceptance. The other is to accept that *p* if it is true that *p*. To ensure logically that incorrigible beliefs attain the objectives of acceptance, we should define incorrigibility as follows:

> The belief that *p* is incorrigible for *S* if and only if (i) it is logically necessary that if *S* believes that *p*, then it is true that *p* and (ii) it is logically necessary that if it is true that *p*, then *S* believes that *p*.

Incorrigible beliefs so defined fulfill the objectives of acceptance as a matter of logical necessity. The first condition might be called the *infallibility* condition because it requires that one cannot fail to attain truth in what one believes, and the second condition might be called the *irresistibility* condition because it requires that one cannot resist believing what is true. Beliefs that fulfill the infallibility condition will be said to be *infallible* beliefs, whereas those that fulfill the irresistibility condition will be said to be *irresistible* beliefs.

This definition of incorrigibility is equivalent to saying that a belief that *p* is incorrigible for *S* just in case (i) it is logically impossible that *S* believe that *p* and that it be false that *p* and (ii) it is logically impossible that it be true that *p* and that *S* not believe that *p*. The first condition, the infallibility condition, is the one that proved insufficient in consideration of necessary truths, but the addition of the second, the irresistibility condition, mends the difficulty. Though it is logically impossible that a person should believe a mathematical truth (for example, that 25 times 26 equals 650) and be in error, it is perfectly possible that a person should fail to believe such a truth. In this case the irresistibility condition is not satisfied, and the belief is not irresistible. Thus, the truth that 25 times 26 equals 650 is not an incorrigible belief for a person when it is logically possible that the person not believe this because it is not an irresistible belief.

Many beliefs that a person has about herself are alleged to be incorrigible in this sense. The favorites are beliefs about conscious mental states of the moment, such as a sharp pain, the idea being that a person cannot be mistaken about what is consciously occurring in her mind at the moment it is occurring. Rather than begin with a discussion of a belief about some mental or psychological state, though, let us go back to Descartes and con-

sider the bare belief of a person that she exists, whatever else might be true about her.

Consider, for example, my belief that I exist. My belief that I exist cannot possibly be false, and it is plausible to affirm that I cannot possibly fail to believe it. Consider next my belief that I believe something. It is logically impossible that I believe that I believe something and do not believe something. The belief is clearly infallible. Moreover, it is plausible to maintain that I cannot fail to believe that I believe something when, in fact, I do believe something and, thus, that the belief is irresistible as well. Thus, the belief that I exist and the belief that I believe something are infallible beliefs and thus plausible candidates for the role of incorrigible beliefs as so defined.

Infallible Beliefs About Thoughts

The foregoing example concerning beliefs about believing something must be distinguished from a closely related one. Once it is conceded that the belief that one believes something is infallible, it might be inferred that if a person believed that he believes that so and so, something of a specific content, then his belief that he believes that so and so is also infallible. This is doubtful, however. I cannot both believe something and be in error in believing that I believe something, but I can believe that I believe some specific thing, that my belief has some specific content, and be in error. The sort of belief that concerns us is acceptance in the interests of obtaining truth and avoiding error, and, as noted in the last chapter, such belief is a functional state implying a readiness to infer and act in specified ways. The inferences and actions of a person may reveal that he does not accept what he sincerely says and even believes he accepts. If a man says he believes a woman is capable of performing a job as well as the man she has replaced and yet immediately infers, without investigation, that everything that goes wrong in the office is her fault, he does not really believe she is as capable as the man she replaced. His chauvinism shows that he does not accept what he says, even if he believes he does. One could offer similar arguments to show that it is logically possible for a person to be mistaken about what she hopes, fears, and wishes.

Are any mental occurrences the objects of infallible beliefs? The best candidates are thoughts and sensations. Let us consider thoughts first. We sometimes say of a person that he thinks that so and so when we are using the term 'think' to mean something very much like belief. That is not the sense of the term we shall consider now. Instead, consider the participial use of the term 'thinking,' which describes an occurrent episode, for example, thinking that Mary is a colonel. Here we use the term to refer to the thoughts that are now occurring to us or our ongoing mental processes. Can a person be mistaken in his beliefs about such occurrences?

Suppose I am thinking that Bacon is the author of *Hamlet*. Suppose, secondly, that I believe that Bacon is identical with Shakespeare, that is, that the man known to us as the author, Shakespeare, is none other than Bacon. However, though I believe this identity to hold, let us also imagine that this belief is not before my mind, I am not thinking of this identity at the time at which I am thinking that Bacon is the author of *Hamlet*. Now, suppose I am asked what I was thinking. I might conclude that I was thinking that Shakespeare was the author of *Hamlet* because, believing that Bacon is Shakespeare, I also believe that thinking that Bacon is such and such is the same thing as thinking that Shakespeare is such and such. Am I correct?

The answer is no. Suppose my thinking, in this instance, consists of my talking to myself, of mulling things over in silent soliloquy, though we admit that not all thinking consists of such silent soliloquy. Nevertheless, there are some cases in which thinking consists of talking to oneself, and focusing on such cases enables us to reveal the way in which a person can be mistaken about what he is thinking. Suppose that when I was thinking that Bacon is the author of *Hamlet,* my thinking consisted of saying to myself: 'Bacon is the author of *Hamlet.*' Now, it is perfectly clear that to say 'Bacon is the author of *Hamlet*' is one thing, and to say 'Shakespeare is the author of *Hamlet,*' another. Thus, thinking that Bacon is the author of *Hamlet* is not necessarily the same thing as thinking that Shakespeare is the author of *Hamlet*. We may, therefore, imagine that I was not thinking the latter when I was thinking the former. Thus, when I reported that I was thinking that Shakespeare is the author of *Hamlet,* and believed what I said, I was quite mistaken. Hence, believing that one is thinking such and such does not logically imply that one is thinking that.[9]

There are several objections to this line of thought that must be met. The first is that even though my thinking that Bacon is the author of *Hamlet* might, in some way, consist of my saying, 'Bacon is the author of *Hamlet*' to myself, it still does not follow that I was not thinking that Shakespeare is the author of *Hamlet* when saying to myself 'Bacon is the author of *Hamlet.*' Maybe I was thinking that Shakespeare is the author of *Hamlet,* but my so thinking did not consist of my saying 'Shakespeare is the author of *Hamlet*' to myself.

The reply to this objection is that there is no reason to say I was thinking Shakespeare was the author of *Hamlet* when I was saying something quite different to myself. My reason for saying that I was thinking Shakespeare is the author of *Hamlet* is a faulty inference. From the premise that Bacon is Shakespeare I inferred that thinking Bacon is such and such is the same thing as thinking Shakespeare is such and such. The inference is as faulty as the inference from that premise to the conclusion that saying Bacon is such and such is the same thing as saying Shakespeare is such

and such. This inference is incorrect, even if the two men are identical. When I said 'Bacon is the author of *Hamlet*,' I did not say 'Shakespeare is the author of *Hamlet*,' even if I think I did at the very time I said what I did.

The argument just enunciated may be obscured by the consideration that if Bacon and Shakespeare are the same, then what I said of the one man is true if and only if what I said of the other is true. However, we avoid this issue by assuming my belief that Bacon is Shakespeare to be false. Indeed, most scholars of Elizabethan literature do assume this. In that case, when I am saying that Shakespeare is the author of *Hamlet,* what I am saying is true, whereas when I am saying that Bacon is the author of *Hamlet,* what I am saying is false. Hence, saying the one thing cannot be identical with saying the other. The same holds for thinking. If, when I am thinking that Shakespeare is the author of *Hamlet,* what I am thinking is true, whereas when I am thinking that Bacon is the author of *Hamlet,* what I am thinking is false—then my thinking the first cannot be identical with my thinking the second. Hence, if I believe that I am thinking Shakespeare is the author of *Hamlet* because I believe I am thinking Bacon is the author of *Hamlet,* as was the case in the example cited, the former belief may be mistaken, even though the latter is correct.

Moreover, I can falsely believe that I am thinking Shakespeare is the author of *Hamlet* at the very same time I am actually thinking that Bacon is the author of *Hamlet.* To see that this is so, notice first of all that I can believe that I am thinking that Bacon is the author of *Hamlet* at the very time at which I am thinking that. My believing that I am thinking can coexist with the thinking and yet be quite distinct from the thinking. When a person talks to herself, she need not believe what she says. If my belief that I am thinking that Bacon is the author of *Hamlet* can exist at the same time as my thinking that, then obviously my false belief that I am thinking Shakespeare is the author of *Hamlet* can exist at the same time as my thinking Bacon is the author of *Hamlet.*

From the preceding argument we may conclude that a person can make all sorts of mistakes about what is presently going on in her mind. The preceding argument may be adapted to show that when a person believes that she is surmising that *p,* doubting that *p,* or pondering that *p,* she may be mistaken in her belief. It would be the most unforgivable pedantry to rerun the preceding argument for each of these states. Moreover, any mental state that has a specific content, the content that *p,* as an object is a state about which one can be mistaken. This should be clear from the preceding argument. Thus, we have subverted the pretentious claim of introspection to be the source of infallible belief concerning the content of even our present thoughts.

Fallible Beliefs About Sensations: Incorrigibility and a Counterexample

Are beliefs about sensations infallible? Just as beliefs about our present thoughts are fallible, so are our beliefs about our present sensations, and for a similar reason. Such beliefs may be based on an inference from false assumptions about the identity of sensations. A person might believe that one sensation is the same as another when this belief is erroneous and, consequently, that he is having one sensation when he is having quite a different one. Let us consider an example of confusing two sensations: hurting and itching. Imagine that a not very sophisticated man goes to his doctor and is inclined to believe what the doctor says. The doctor tells the man that it is not surprising that his sensation is sometimes one of pain and sometimes one of itching because itches are really pains. All itches, she says, are pains, though some are very mild. The explanation for this, the doctors affirms, is that the nerves that lead to pains and itches are the same, and whether one feels a pain or an itch depends on the intensity of the stimulation of the nerve endings. Itches are the result of less intense stimulations but, medically considered, itches are pains and the same kind of state. There is no difference between them other than the intensity of the stimulation of the nerve endings. Such is the authority of the doctor, and such the credulity of the man, that her word is taken as creed. From that moment on, he never doubts that itches are pains, and, though they feel different, he firmly believes that he is in pain, even if only very slightly so, whenever he has the slightest itch. When he itches, therefore, he erroneously believes that he is in pain, even if only very slightly so. Thus, his beliefs that he is in pain are often erroneous and are by no means infallible.

It should be apparent that the man in question might have been misled by his esteemed medical sage into believing that one sensation was another when in fact it was not. He might, without understanding how such things could possibly be true, believe that they are. Hence, beliefs about sensations, like beliefs about thoughts, are fallible and corrigible. In general, very little of what we believe about our own mental and psychological states is incorrigible. Error can, as a matter of logic, insert itself stealthily between belief and what is believed in this matter as in others.

Other Alleged Counterexamples

Other counterexamples to the thesis that beliefs about sensations are incorrigible concern people who are in a more or less aberrant state. Consider a man who believes he is about to undergo some painful experience, for example, that he is about to be touched by some very hot object, though in fact the object is cold to the touch. Because he expects to feel a burning sen-

sation, he could believe that he is feeling such a sensation during the first moment or two that the cold object is touching his flesh. This belief would then be false. The difficulty with such a counterexample is that it is problematic whether expecting to feel a burning sensation produces false beliefs or a burning sensation.

Another kind of counterexample concerns those who are mentally aberrant and experiencing hallucinations. For example, suppose that a very paranoid man complains that he is suffering excruciating pain because little green Martians are cutting into his flesh. If he then goes on to tell us that the reason he shows no sign of being in pain (he does not grimace, wince, and so forth) is that he does not want to let the Martians know they are succeeding in making him suffer, we might begin to think he believes he is in pain when he is not. The man might in fact not be undergoing any pain whatsoever, even though he does genuinely believe that he is suffering. He might later, when the aberrance has vanished, report that this is what had happened. Some philosophers might doubt that the usual concepts of belief and sensation apply in such peculiar cases as this, and others would have other doubts. Such examples may be genuine counterexamples, however.

Fallibility and Inference: Summary of the Argument

The arguments in regard to the fallibility of beliefs concerning thought and sensation may be summarized as follows. Whatever one can believe as a result of introspection, one can instead believe as a result of inference, and the inference can be based on false premises. If a woman believes she is in pain as a result of feeling pain, then, of course, she will be correct in believing that she is in pain. If, however, a man believes that he is in pain because some scientific or religious authority figure tells him that he is in pain, then what he believes may well be false. In such a case, the person has accepted a premise, namely, that what the scientific or religious authority figure says is true which, together with the premise that the authority says the person is in pain, leads the person to infer and, therefore, to believe that he or she is in pain and to believe this falsely when the authority is untruthful.

Inference from the testimony of an authority is only one example of how false beliefs about one's mental states may result from inference. It may seem strange to imagine people coming to believe they have some thought or sensation as a result of inference from testimony, but this is the strangeness of the human mind, not of logic. People may believe things in ways that are quite unreasonable and mentally deranged. As a result of these mad beliefs, they may come to believe things about anything, about their very own thoughts and sensations, which they would otherwise never believe.

Nomological Infallibility

The preceding observation may be extended to refute attempts to salvage the doctrine of infallibility by replacing the notion of logical possibility with some weaker notion of possibility in the definition of incorrigibility. One might, for example, attempt to rescue the doctrine of incorrigibility by substituting a notion of *nomological necessity* or impossibility, that is, necessity or impossibility in terms of the laws of nature, for the notion of logical impossibility in the definition of incorrigibility. This would have a beneficial consequence concerning the second condition of incorrigibility, the irresistibility condition. It is much more plausible to suppose that it is the result of a law of nature, or a law of psychology, that a person believes that she is thinking or feeling something when she is, than to suppose that this is a consequence of logical necessity.

The amendment to nomological necessity or impossibility will be to no avail for rescuing the doctrine of infallible beliefs, however. It is logically possible to infer false conclusions about one's thoughts and sensations from false beliefs and also in terms of the laws of human nature. There are, for example, people belonging to religious groups who believe that pain is unreal and, therefore, that they do not have pain, though they clearly suffer like the rest of us when injured. Some paranoid people believe they are in pain, as noted above, when they believe they are attacked by powerful enemies. It is obvious, however, that they suffer no physical pain. The human mind provides us with no prophylactic against error, even concerning our own thoughts and sensations, as the strange beliefs of humanity, arising from hopes and fears, abundantly illustrate. There is nothing so foolish that we cannot believe it if it is repeated often enough and with enough authority. Every demagogue understands this very well.

Meaning and Belief

There is a familiar but erroneous objection against the preceding line of thought. It is that the people who hold odd beliefs really have different beliefs than they appear to have. The words they utter have a meaning that differs from what is customary. On this account, the man misled by his doctor does not believe he is in pain when he itches; he just attaches a different meaning to the word 'pain' so that it means 'pain or itch' and has no false beliefs about pains, though he appears to do so. Similarly, the paranoid person means something different by 'pain,' as do the members of the religious cult who think there is no pain. How should we reply to this objection?

Current theories of meaning are fraught with controversy, and so it will not be possible to reply definitively by appeal to any such theory. The problem is rendered yet more difficult by argumentation, primarily from

Quine,[10] that there is no sharp boundary between what we explain in terms of change of meaning and what we explain in terms of change of belief. Finally, a methodological principle advanced by Davidson,[11] the principle of *charity,* exhorts us to interpret the utterances of others in such a way as to render the beliefs we ascribe to them as true as far as is charitably possible. These considerations might appear to support the objection in that, given the lack of a sharp boundary between meaning and belief, the charitable thing would be to interpret the odd remarks of the people in our examples in such a way as to ascribe true rather than false beliefs to them. To do this, we would need to suppose that the man who says he is in pain when he itches means something different by 'pain' than is customary.

The reply is simple and appeals to simplicity. We have a simple explanation as to why people say what they do that does not require the complexity involved in supposing that they have changed the meaning of the words they utter from what is customary. The simple explanation is that they acquired an odd belief, one they affirm, and that they mean what they say. The application of the principle of *charity* would, consequently, be misapplied in such cases if we were to assume they meant something different from what is customary.

Let us consider the example of the person who has come to believe that itches are mild pains. There is something in his speech behavior that might lead us to consider the hypothesis that his meaning of the word 'pain' is deviant and, more specifically, that he means by the word 'pain' what we mean by the expression 'pain or itch.' He calls itches pains, after all, and this favors the hypothesis that his meaning is deviant. But there is evidence against this hypothesis as well. His linguistic training is altogether ordinary and like the rest of ours. This fact conflicts with the hypothesis that his meaning of 'pain' is deviant and raises the question, Why does he say he is experiencing pain when he itches?

One hypothesis is that he is physiologically peculiar and he hurts when we itch, but, upon questioning him, we learn the correct answer. The doctor has convinced him that itches are pains, very mild ones. He tells us that this is a very peculiar fact about itches discovered by medical research. In the light of this evidence, the simplest explanation for all his behavior is that his meaning of 'pain' is standard, but one of his beliefs about pains is peculiar. He believes that itches are pains. The crucial reason is this. The hypothesis that he means by 'pain' what we mean by 'pain or itch' fails to explain what he tells us about what he learned from the doctor. For this reason, the hypothesis of meaning change must be rejected. If by 'pain' he meant 'pain or itch,' then he would not consider it to be peculiar that itches are pains. That would amount to the simple tautology that itches are either pains or itches, and there is nothing peculiar about that. The simplest hy-

pothesis to account for the total data obtained from the man, including the fact that he too thinks it is odd that itches are pains, is that his meaning of the word 'pain' is standard but one of his beliefs about pains is deviant.

The foregoing remarks apply with even greater clarity to the other examples of false beliefs about one's own sensations. It would be uncharitable to accept an interpretation of the meaning of what the people in the counterexamples say that would assign deviant meaning to their words in order to ascribe true beliefs about their sensations to them. To do so would render their exotically different general beliefs about the world mere tautologies. That would fail to explain why they think these beliefs are profound facts about the nature of things. It might be possible to invent some hypothesis compatible with the deviant meaning hypothesis to account for this oddity. The simplest hypothesis, however, is that the meanings of the words that people utter are standard and the beliefs they have acquired are peculiar. That is what they say, after all, so why not believe them? What could be simpler?

We may now conclude our remarks on the incorrigible. We have found almost no beliefs about contingent matters that are incorrigible in the relevant sense of the logical or nomological impossibility of error. The belief that I exist and the belief that I believe are infallible, but any belief about what I think or believe about any feeling or sensation, as well as other contingent matters, is fallible and subject to correction. Moreover, even these infallible beliefs may not be irresistible and hence not incorrigible. There may some extreme materialists who do not accept that they or others *believe* because believing is not a material state and is, therefore, rejected. There may be some extreme nihilists who do not accept that they or others exist because existence is rejected. Incorrigible beliefs are, therefore, inadequate to justify those beliefs we consider well enough justified to constitute knowledge. We must abandon the quest for incorrigible foundations or embrace the skeptical result.

The Opacity of Incorrigibility

There is another aspect of the features of incorrigibility or infallibility that makes them inappropriate as the basis of justification, namely, their intellectual opacity. A belief may be infallible even though the person has no idea that this is so. When the infallibility of a belief is concealed from the subject of the belief, it is powerless to justify the belief for the person. The infallibility of the belief may not reveal itself to the subject of the belief; it may remain concealed and opaque rather than revealed and transparent. Opaque infallibility is too dark a feature to yield the light of evidence. We noted the example of a belief that was infallible but not justified at all, namely, a belief in some mathematical proposition which, though it hap-

pened to be true, was based on a false assumption and fallacious reasoning. The example illustrates the problem of the opacity of infallibility.

Recall the person who believed that there was a one-to-one correspondence between the set of whole numbers and the set of odd numbers because he believed that after a certain point, there were no more whole numbers. The reasoning would be absurd, but, since it is a necessary truth that there is a one-to-one correspondence between the set of whole numbers and the set of odd numbers, it is impossible that anyone believing this should believe something false. It is impossible, logically impossible, that a person should believe something that is necessarily true and be in error, for it is logically impossible that anything necessarily true should be false.

The problem is that, though the proposition is necessarily true, the necessity is not understood by the person who believes the proposition, it is not something he sees, and thus the necessity remains opaque. When the necessity is opaque to the person who believes it, then the person may, as in the case considered, fail to be justified. This is a problem that we shall encounter on other accounts of justification, not just that of the infallible foundationalist, for it arises when some feature is alleged to generate justification even when the presence of the feature is opaque to the believer.

Moreover, even if the proposition believed is necessarily true, the necessity may remain opaque. For example, consider the proposition that someone believes something. If someone believes that proposition, her belief is infallible. It is a necessary truth that *if* a person believes that someone believes something, *then* someone believes something. But the person believing this may easily fail to notice the necessity. Some philosophers and others interested in the subject of infallibility may notice the necessity of the hypothetical claim and become justified on the basis of it. For the rest of those who believe that someone believes something, the necessity will remain opaque and fail to generate justification.

There is a final objection to the claim that infallibility might be the basis of justification. The notion is implicitly one concerning the necessity of a hypothetical. It says that a belief that p is infallible if and only if it is logically impossible that someone should believe that p and that p should not be true. This is equivalent to the claim that it is necessarily true that *if* a person believes that p, *then* it is true that p. Once the hypothetical feature is exposed, it becomes clear that one could concede the necessity of the hypothetical and yet deny the justification of the belief on the grounds that the antecedent of the hypothetical is not satisfied.

Consider the claim, I think, therefore I exist, taken from Descartes. It is necessarily true that if I think, then I exist, but someone wishing to deny that I am justified in the claim that I think or that I exist might deny the truth of the antecedent, that is, of the hypothetical *if* I think. Descartes assumed that he doubted and therefore that he thought and existed. One

might agree that *if* Descartes doubted, then he thought and existed, for these are necessarily true consequences, but avoid the consequences by denying the antecedent, the *if* I think. An extreme materialist might reject Descartes' assumption that he doubts, that he thinks, and avoid the consequence. A philosopher who rejects the existence of the self might reject the existence of a thinker while allowing the existence of a thought. There are other possibilities. The point is that to arrive at the conclusion in the first person that I am justified in believing something, even self-justified, one needs the assumption that the person believes something, and that too might be controverted. The conclusion is that even a philosopher who agrees that some beliefs would be infallible (if there are such beliefs) may reject the conclusion that the beliefs are justified, even self-justified, by rejecting the claim that there are such beliefs.

Justification As a Logical Guarantee of Truth for Nonbasic Beliefs

Just as logical guarantees have been traditionally sought for the truth of basic beliefs by the infallible foundationalist, so it has been thought that basic beliefs must guarantee the truth of nonbasic beliefs. Though this sort of theory appears unable to provide an adequate supply of infallible basic beliefs, it is worthwhile to consider the question of whether, assuming a larger supply of basic beliefs, they could be expected to guarantee the truth of the nonbasic beliefs we take to be justified in accepting. There are two reasons for considering this. First, of course, we must allow for our own fallibility in argumentation. Perhaps there are more infallible basic beliefs than we have dreamt of in our philosophy. Second, and more important, some supply of fallible basic beliefs might, if true, guarantee the truth of all nonbasic beliefs we are justified in accepting. In that case, a mixed foundation theory allowing that basic beliefs be fallible but requiring that the justification of nonbasic beliefs must guarantee their truth would be acceptable.

We shall, therefore, now consider the attempt to provide a theory of the justification of nonbasic beliefs on the basis of basic beliefs that logically guarantee the truth of the nonbasic beliefs. This attempt has often led to some sort of analytically reductive theory affirming that the content of nonbasic beliefs can be reduced by logical analysis to the content of basic beliefs. One such theory, typical of analytically reductive theories, is *phenomenalism,* a modern analytic refinement of the theory of Bishop Berkeley.[12]

Phenomenalism

Berkeley held that we have immediate knowledge of our own ideas, which include the appearances of sense. Suppose beliefs concerning appearances

are basic beliefs. In the opening sections of the *Principles,* Berkeley suggests that what we mean when we affirm the existence of external objects, of a tomato, for example, can be expressed in terms of what ideas we have of it, what appearances we experience (i.e., the reddish, roundish, bulgy ones), and of the appearances we would experience, if we were to undertake various courses of action. For example, we would experience squishy, wet, runny appearances if we were to undertake to strike out toward the reddish, roundish, bulgy appearances.[13] Thus, the contents of our beliefs about physical objects or the statements expressing those contents would be equivalent in meaning on this view to statements about what ideas we have, what appearances we experience, conjoined to hypothetical statements about what ideas we would have and what experiences we would experience if various conditions were fulfilled.

In modern dress, this is the doctrine of *phenomenalism,* which affirms that all statements and the contents of all beliefs concerning external objects can be translated without loss of meaning into statements about what appearances we experience or would experience under various conditions; or, to adopt the terminology of sense data in place of that of appearance, about what sense data we experience or would experience under various conditions.[14] The plausibility of the doctrine can be appreciated by reflecting on a thought experiment. Suppose that all experiences that you have or would have under any conditions are just what they would be if the external world existed but the external world has, in fact, vanished. The absence of the external world would be completely beyond detection. Now that you have imagined this, imagine that this is what, in fact, has occurred. You have no indication of your loss, of course, because all that you experience is just what you would experience if the external world had not vanished.

If you feel that this thought experiment is a verbal or semantic trick, that nothing has vanished in it, then you will understand the motivation for accepting the doctrine of phenomenalism. Phenomenalism tells us that if all the sense data we experience and would experience under any conditions are exactly the same as if the external world were to exist, then the external world *does* exist. To say that it exists just means that we do and would experience the appropriate sense data under various conditions. The imagined disappearance of the external world is a semantic illusion.

The relevance of the doctrine of phenomenalism to the foundation theory is that, if phenomenalism is true, then conjunctions of statements about what sense data we experience or would experience under various conditions logically imply statements about the external world because they would exhaust the meaning of those statements. So conjunctions of statements about sense data would, if true, logically guarantee the truth of the statements about the external world. Put in another way, phenomenalism tells us that statements about the external world are analytically reducible

to statements about sense data, and therefore the truth of the latter guarantee the truth of the former. Assuming sense data statements to express the contents of basic beliefs, their truth would guarantee the truth of nonbasic beliefs about external objects. Finally, therefore, if the truth of the basic beliefs is self-guaranteed, that guarantee would extend to the nonbasic beliefs about external objects that the basic beliefs logically imply as a result of the reduction of the former to the latter.

Phenomenalism, in the analytic version, analytic phenomenalism, though not commonly defended currently, is characteristic of analytically reductive theories and exhibits the relevance of such theories to the foundation theory of justification. Other such analytically reductive theories include analytic behaviorism, which holds that all statements about mental states, thoughts, feeling, and sensation, are reducible in meaning to statements describing behavior, and analytic empiricism, which holds that all statements about theoretical entities, atoms, quanta, and microwaves, are reducible in meaning to statements describing observables. All such reductive theories allege that some kind of statements, the targets of the reduction, are equivalent in meaning and, therefore, logically equivalent to the statements to which they are reduced. If the latter statements express the contents of basic beliefs, then the reduction explains how they can guarantee the truth of the statements that are the target of the reduction. They logically imply the target statements and, assuming the truth of the basic beliefs to be guaranteed, would logically guarantee the truth of the nonbasic beliefs.

Objections to Phenomenalism

There are, however, a number of problems concerning analytic phenomenalism that have led to its rejection; similar problems have led to the rejection of other reductive theories, though the issue of reductionism remains controversial for other reductive theories. It will, therefore, be useful to consider the difficulties confronting phenomenalism as representative of the sort of difficulties confronting reductive theories, though, of course, not all the difficulties of phenomenalism confront other reductive theories. Here then are the difficulties confronting phenomenalism. First, the reductive language, the language of sense data that describes the appearances one senses or the way in which one is appeared to, leads to controversy because of the claim of the privileged status of reductive language. For example, Ayer suggested that some sense-data statements are incorrigible,[15] and we have considered the problems surrounding such a claim. Second, some philosophers doubt that the required meaning analysis of statements about external objects in terms of sense-data statements can be effected. Chisholm, for example, has argued convincingly that no statement about an external object logically implies any statement about sense data or ap-

pearances and, therefore, that the meaning equivalence fails.[16] It is possible, nevertheless, that the meaning of statements about external objects is exhausted by the meaning of statements about sense data in the sense that the latter logically imply the former, though not vice versa. The exhibition of such logical implications would yield an important kind of reduction for foundationalism. If statements about sense data, even if not incorrigible, logically imply statements about external objects, even if not vice versa, then basic beliefs about sense data, if true, can logically guarantee the truth of nonbasic beliefs about external objects.

The appeal to analytically reductive theories to support foundationalism faces a decisive problem that can be illustrated by further consideration of phenomenalism. Suppose we have a statement about an external object, *E* (for example, that there is a tomato in front of me) and we have a phenomenalistic analysis of *E* in sense-data language, consisting of a conjunction of sense-data statements *S1, S2,* and so forth, through *Sn.* For the sake of simplicity, let us refer to the statements expressing the contents of basic beliefs as basic statements and those statements expressing the content of nonbasic beliefs as nonbasic statements. Consider the statements *S1, S2,* and so forth, to *Sn.* Are these statements basic or nonbasic ones?

Some of the sense-data statements *S1,* S2, and so forth, to *Sn* that analyze or even logically imply the external object statement *E* must be nonbasic. We can illustrate this by appeal to the statement that there is a ripe tomato before me. Consider the sort of sense-data statements one might think are part of the analysis of this statement. Some of these statements would be about what I am sensing at the moment, for example, a reddish, roundish sense datum, and these might be basic, but they are not logically sufficient to analyze or imply logically that there is a ripe tomato before me. We would also require hypothetical, indeed, contrary-to-fact hypothetical, contingent statements about what I would be sensing if I were to alter the circumstances, for example, by striking in the direction of the sense data with the intention of squishing the tomato. If such efforts produce no alteration, the sense data may be deceptive, the stuff of dreams and hallucinations, rather than those of a genuine ripe tomato.

In short, the hypothetical statements in question must articulate what sense data I would sense under various conditions if there were a ripe tomato before me, in order to yield the conclusion that there is a ripe tomato before me. Some of these hypothetical statements must be nonbasic, because they would have to be justified by evidence, if they are justified at all. Many of them would be contrary to fact, asserting what one would sense if certain facts were other than they are. Belief in the truth of a contrary-to-fact hypothetical contingent statement, if it is justified at all, is justified on the basis of evidence. Therefore, the set of statements *S1, S2,* and so forth, to *Sn* of any plausible phenomenalistic analysis or reduction can-

not all be self-justified basic beliefs about sense data. The upshot of this argument is that the sense-data statements *S1, S2,* and so forth, to *Sn* of any plausible phenomenalistic analysis or reduction of a statement *E* about an external object cannot all be basic. If not all the sense-data statements are basic, then the analysis or reduction does not provide us with a set of basic statements that guarantee the truth of a nonbasic statement.

The preceding argument can be extended to a variety of analytically reductive theories. Philosophers who have eschewed phenomenalism as unrealistic have often embraced some other analytically reductive theory to sustain their own version of a foundation theory. For example, some philosophers of science have regarded observation statements as basic and have proposed some reductive analysis of generalizations and theories in terms of observation statements. It was once argued, for instance, that generalizations of the form 'Anything that is *O1* is *O2*,' where '*O1*' and '*O2*' are observation terms, may be analyzed as a conjunctive statement: If *x1* is *O1,* then *x1* is *O2,* and if *x2* is *O1,* then *x2* is *O2,* and so forth. Here the difficulty mentioned above becomes obvious. Since it is clear that not all the hypothetical statements are self-justified, at least some of them must be justified, if they are justified at all, as nonbasic beliefs, that is, their justification must be based on evidence. The reason is that we shall not have observed every one of the objects *x1, x2,* and so forth, and, thus, even if we allow that beliefs in categorical observation statements are self-justified, not all the hypothetical statements in the analysis of the generalization are self-justified. Consequently, some of those beliefs will be nonbasic. Hence, such a reductive analysis will not show how basic observation statements guarantee the truth of nonbasic generalizations.

Similar remarks apply to reductive analyses of theoretical statements in terms of observation statements. Thus, reductive analysis, though motivated by foundationalism, fails to support it because the reduction will leave us with a base of hypothetical statements in the preferred vocabulary of sense data or observation. These hypothetical statements will not, however, supply us with a foundation. We will be justified in accepting them, if we are justified at all, only on the basis of evidence, and therefore they are nonbasic. The objective of reduction is to reduce nonbasic statements to a collection or conjunction of basic statements. Reduction fails in this objective exactly because it leaves us with a set of different but equally nonbasic statements.

Summary

We have seen that the quest for infallible foundations is a failure. The attempt to find infallible basic beliefs that guarantee their own truth to serve as a foundation yielded the most meager results. Fallibility infects almost

all our beliefs. More important, however, is the fact that the infallibility of infallible beliefs may be opaque to the subject so that the subject has no idea that the beliefs are infallible. Infallibility is a feature of a belief that is itself nonbasic, that is, the belief that a belief is infallible must be justified by argument or, at least, by evidence. Consequently, the infallibility of a belief does not ensure that the person is justified in the belief and, as we noticed in the case of a belief in a mathematical truth arrived at by fallacious reasoning, the person may be unjustified in belief even though the belief is infallible. Infallibility of a belief may fail to provide any justification for a belief to a person who is ignorant of the infallibility of it. The person cannot see through the opacity of the infallibility of the belief to the justification of the belief. The infallible feature of the belief is concealed from the subject and is impotent to provide the light of evidence for the justification of her belief.

The attempt to extend the guarantee of truth from basic to nonbasic beliefs by undertaking to reduce the content of the latter to the contents of collections or conjunctions of the former is equally unsuccessful. The reduction leaves us with different but equally nonbasic beliefs. The idea that we might construct or reconstruct the edifice of knowledge from a set of basic beliefs whose truth is guaranteed and that guarantee the truth of all the rest was of extraordinary importance in the theory of knowledge. Had it been successful, it would have provided us with a means of ensuring the truth of what we accept and avoiding the problems arising from the existence of justified false beliefs. Like other philosophical traditions, it taught us something different from what was originally intended. The lesson is that we are fallible in what we believe and must proceed without any guarantee of our success. The quest for truth, if based on a foundation of self-justified beliefs, must be based on a fallible foundation.

Introduction to the Literature

There is an excellent discussion of foundationalism in John Pollock and Joseph Cruz's *Contemporary Theories of Knowledge,* including, in chapter 7, a defense of a kind of direct realism that resembles foundationalism, although the foundational states are not beliefs. This view is an alternative to the kind of foundationalism discussed in this chapter. The traditional forms of foundationalism are exposited and defended by Panayot Butchvarov in *The Concept of Knowledge;* by Arthur Danto in *Analytical Philosophy of Knowledge;* by Bertrand Russell in *The Problems of Philosophy;* and by A. J. Ayer in *Foundations of Empirical Knowledge.* Roderick Chisholm defended something akin to an infallible foundation theory in his early work, *Perceiving: A Philosophical Study.* See also Carl Ginet's *Knowledge, Perception, and Memory;* and Paul Moser's *Empirical Justification.*

Notes

1. Fallible foundations are defended in John L. Pollock and Joseph Cruz, *Contemporary Theories of Knowledge* (Oxford, N.J.: Rowman & Littlefield, 1999), chap. 5; and in R. M. Chisholm, *The Foundations of Knowledge* (Minneapolis: University of Minnesota Press, 1982). See discussion in chapter 1 n. 13.

2. For example, see R. M. Chisholm, *Theory of Knowledge* (Englewood Cliffs, N.J.: Prentice-Hall, 1966), chap. 2, "The Directly Evident"; and *Perceiving: A Philosophical Study* (Ithaca, N.Y.: Cornell University Press, 1957), particularly chap. 5, "Justification and Perception."

3. Pollock and Cruz have suggested this; see *Contemporary Theories of Knowledge,* chap. 2. See also Jonathan Kvanvig, "How to Be a Reliabilist," *American Philosophical Quarterly* 23 (1986).

4. See Laurence BonJour, *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press, 1985), particularly chaps. 1, 2, 4.

5. Descartes is often claimed to be such a rationalist. Historically, this position is probably mistaken. In a late section of the *Discourse on Method,* Descartes mentions how he had recourse to experiment to decide between competing hypotheses. See *Discourse on Method,* pt. 6. Thus, it is doubtful that Descartes was a strict rationalist. He seems to have agreed that at least on some occasions justification is derived from sense experience.

6. Panayot Butchvarov in *The Concept of Knowledge* (Evanston, Ill.: Northwestern University Press, 1970); and Arthur C. Danto in *Analytical Philosophy of Knowledge* (Cambridge: Cambridge University Press, 1968) have expounded this view. See Butchvarov's discussion of the notion of sufficient evidence on pp. 49–50, and Danto's discussions of both direct knowledge and adequate evidence on pp. 26–49, 147, and 122, respectively. A. J. Ayer in *Foundations of Empirical Knowledge* (New York: St. Martin's, 1955), 74–84, proceeds with the same assumptions and attempts to found justification on incorrigible statements.

7. J. L. Pollock, *Knowledge and Justification* (Princeton: Princeton University Press, 1974); and with Joseph Cruz, *Contemporary Theories of Knowledge;* Mark Pastin, "Modest Foundationalism and Self-Warrant," *American Philosophical Quarterly,* Monograph Series 9 (1975); Ernest Sosa, "Epistemic Presupposition," in *Justification and Knowledge,* ed. G. S. Pappas (Dordrecht: Reidel, 1979), 79–92; Richard Foley, *A Theory of Epistemic Rationality* (Cambridge: Harvard University Press, 1987); and William Alston, "Two Types of Foundationalism," *Journal of Philosophy* 73 (1976): 165–85.

8. Another method for dealing with these problems is presented by George Nakhnikian in "Incorrigibility," *Philosophical Quarterly* 28 (1968): 207–15, but Nakhnikian defines incorrigibility in terms of knowledge, which would be unacceptable in the present context.

9. Katherine Pyne Parsons, "Mistaking Sensations," *Philosophical Review* 79 (1970): 201–13, raises some similar objections to the doctrine that statements about mental states are incorrigible.

10. W. V. O. Quine, *Word and Object* (Cambridge: MIT Press, 1960).

11. Donald Davidson, *Inquiries into Truth and Interpretation* (Oxford: Clarendon, 1984).

12. George Berkeley, *A Treatise Concerning the Principles of Human Knowledge*, ed. Colin M. Turbayne (Indianapolis: Bobbs-Merrill, 1957), pt. 1.

13. Ibid.

14. Ayer, *Foundations*, 74–84.

15. See Ayer's defense of this thesis in *Foundations*, 74–84; and in "Basic Propositions," in *Philosophical Essays* (New York: St. Martin's, 1954), 105–24, esp. 113ff.

16. See Chisholm, *Perceiving*, appendix, 190–93.

4

Fallible Foundations

THE PRECEDING CHAPTER has shown why a foundation theory of justification must subscribe to the doctrine that at least some basic beliefs are fallible, or else embrace skepticism. The number of infallible beliefs is far too restricted to support our commonsense claims to knowledge. Foundationalism is the doctrine that self-justified beliefs constitute the foundation of knowledge. Can the edifice of knowledge be based on a foundation without a guarantee of truth? Is it tenable for the foundation theorist to allow that beliefs which are fallible, which may be false, are nevertheless basic and self-justified beliefs?

Let us consider the merits and shortcomings of a fallible foundationalism advocating that knowledge rests on fallible but self-justified beliefs. Thomas Reid claimed that some beliefs, for example, perceptual beliefs concerning what we see clearly and distinctly before us, are justified in themselves without need of supporting arguments, even though we have no guarantee that they are true. In short, such beliefs are self-justified because their justification is inherent and is not derived from reasoning. As Reid put it, they are beliefs of common sense that have a right of ancient possession and, until this inherent right is successfully challenged, they remain justified without support from any other beliefs.[1] They possess a kind of self-justification that is fairly regarded as innocent until proven guilty.

Reid hit upon a critically important line of defense for the fallible foundation theory. According to him, some beliefs are worthy of our trust even if we have no guarantee that they are true much as an experienced guide is worthy of our trust, even though we have no guarantee that she will bring us to our destination. Though we remain vigilant to detect errors, we may, in the customary affairs of life, rely on the intrinsic justification that at-

taches as a birthright to various of our beliefs. They are justified in themselves without appeal to independent information. If this doctrine proves tenable, it could provide us with a set of basic but fallible beliefs. The rights of birth and ancient possession have, however, been challenged in the political sphere, and we must bring them under close scrutiny here as well.

Perceptual Belief and Independent Information

Candidates for the status of self-justified beliefs whose justification does not depend on independent information are frequently perceptual. Let us consider some promising candidates and see if we can discover any worthy of the office. Suppose I believe I see a typewriter beneath my fingers. For my belief to be justified, I must have some independent information about what I see, namely, that a thing that looks like the thing beneath my fingers is a typewriter. If I did not have that information, then I would not be justified in believing that I see a typewriter. In short, whenever I see a thing of a certain kind, my being justified in the belief that I am seeing a thing of that kind depends on independent information I have about how things of that kind look. This information justifies me in concluding that the thing I see is such an object.[2]

It is clear that a similar argument will confront us when we consider a simpler belief, for example, the belief that I see something red. It will be argued that, for me to determine that what I see is a red thing, I must have independent information about how red things look and, indeed, about how they look under various conditions. Even if we assume that for a thing to be red is just for it to look red under standard conditions to normal observers, for me to determine that what I see is red, I must know how to tell when conditions are standard and when an observer is normal. Thus, if something looks red to a person, she cannot justifiably conclude that it is red from the formula that red things look red in standard conditions to normal observers. She would also need to know that the conditions are standard and that she is normal. Independent information is, therefore, required for the justification of this perceptual belief. More generally, to be justified in accepting such a belief, one requires information about oneself and the conditions of perception.

The question to examine next is whether any more cautious perceptual belief has a justification that does not depend on independent information. A prime candidate is the belief that I see something, without specifying what sort of thing it is that I see. Here, one might think, is a belief that does not require any independent information for one to be justified in accepting it. Nonetheless, there is reason to doubt this if one construes the word

'something' in such a way as to imply that the object seen is some real thing and not, for example, something hallucinated. Once again, there is need of independent information that would enable the person in question to determine that this is a case of seeing something and not merely a case of hallucination or dreaming, or whatnot. Since one may hallucinate, one cannot justifiably accept that one sees something as opposed to merely hallucinating unless one has information enabling one to distinguish hallucination from the real thing.

Justification and Innocent Belief

There are objections, rather standard ones, to the preceding line of thought. It might be objected, for example, that one does not need to know anything about hallucination in order to accept justifiably that one sees something. A person who never had hallucinatory or any other deceptive experiences might accept with justification that she sees something even though she lacks any information about deceptive experience. If it be asked how she can be so justified in her belief, even though she lacks the information to determine whether she is seeing and not hallucinating, the answer is that such beliefs do not require the support of argument or independent information. They are justified until they are shown to be erroneous or unjustified. They are, as Reid suggested, innocent until proven guilty.[3]

No doubt, in everyday situations we allow uncritically that such beliefs are reasonable until they are shown to be incorrect, but, as noted in the first chapter, a person may be reasonable to accept something that he is not justified in accepting in way that would lead to knowledge. No doubt in the case of the person who was innocent of hallucinatory experience, we would uncritically allow that his perceptual belief was reasonable, but we may deny that the person has knowledge. A little critical circumspection shows that common sense should not be allowed to run unbridled in the epistemic field. All sorts of perceptual beliefs, the belief that one saw a bear print, for example, are considered reasonable when we have no great stake in the question of whether the belief is true or false. However, when a great deal (our personal safety, for example) hinges on the matter of whether the person saw a bear print or something else, then we become instantly more cautious and exacting and require the kind of justification that yields knowledge. We require a park ranger to be able to tell a bear print when she sees one so that when she believes she sees a bear print she is justified in a way that gives her knowledge.

We are casual about conceding the reasonableness of a belief until something of practical importance or epistemic consequence rests on the question. Perceptual beliefs are considered innocent until proven guilty when we care not the least whether the belief is innocent or guilty. Once we do care,

though, we start to ask serious questions, the ones concerning justification. The very first is whether the person is justified in accepting what he does in the interests of obtaining truth and avoiding error on the basis of his information. We seek to determine if the person has information that would enable him to determine whether he actually sees the thing he thinks he does and render him justified. His belief is presumed to be neither justified nor unjustified but is evaluated in terms of the information he possesses. Whether he is justified in accepting what he does or not depends on the adequacy of his information.

An Objection: Reliability or Information?

It might be objected to the preceding that one does not require information to be justified in such a belief but rather a reliable competence. An externalist, who maintains that the justification of a belief is determined solely by the relation of the belief to the external world, might object that the only requirement is that one's belief arise in a reliable manner.[4] To be able to tell whether one is seeing something or not is obviously essential to being justified in accepting that one is seeing something, but, it might be argued, the competence required to tell this need not involve the acquisition of information. One can have the ability to tell whether something is of a certain sort without needing any information to make such determinations. It is sufficient to respond to experience in a reliable manner.

The foregoing objection must be met squarely. A person may learn to tell whether or not he is seeing something without appealing to any premises or making any conscious inferences. Gilbert Harman has suggested we might nevertheless construe such cases as examples of unconscious inference.[5] It suffices, however, to note that a person may be said to have information he cannot easily describe and to employ such information in various ways. For example, suppose I know the shortest route from Rochester to Buffalo, though I cannot tell you the name of the highway. Moreover, imagine that I am not very good at giving directions, so I cannot tell you how to get from Rochester to Buffalo. Does this show that my ability to get from Rochester to Buffalo does not depend on the information I have about the route from Rochester to Buffalo?

Hardly. I obviously do have the information I need to get from Rochester to Buffalo, though I may be very poor at conveying this information to others. That I make the trip successfully on many occasions shows that I have the required information. My reliability depends on my ability to employ information, which I might find difficult to articulate, about the route from Rochester to Buffalo. Similarly, my reliability in accepting that I see when I do, or even in accepting that I feel or think when I do, depends on my abil-

ity to employ information, which I might find difficult to articulate, about seeing, feeling, and thinking.

Any belief whatever is open to similar argument concerning the need for independent information. Indeed, even the very subjective belief that it seems to me that I am seeing something is one I am justified in accepting only if I have the information needed to tell whether it seems to me that I am seeing something or having some quite different experience, for example, that of wondering whether I am seeing something. I may wonder whether I am seeing something when it does not especially seem to me that I am seeing something, and unless I have the information required to tell the difference between wondering and seeming, I am not justified in accepting that I am in one state rather than the other. Even our most modest beliefs turn out to be ones requiring independent information to justify us in accepting them for the purposes of obtaining truth and avoiding error. The preceding argument uncovers a ubiquitous need for independent information to justify acceptance aimed at truth, and in so doing it undermines the foundation theory.

Chisholm and the Noncomparative Use of Words

The most important line of reply on behalf of the foundation theory is provided by R. M. Chisholm, who is responsible for calling our attention to the need for independent information to justify our perceptual beliefs.[6] He contends, however, that some beliefs, those articulated in noncomparative terms, do not require independent information for their justification. He distinguishes between the comparative and noncomparative uses of certain words. Ordinarily, when we apply a word, whether to our own states or to things, the application is based on a comparison we make. For example, if we say something appears red, we may be comparing the way this thing appears with the way other things appear. It is analytic or true by definition in the comparative use of 'appears red' to say that red things appear red to normal observers in daylight. In the comparative use of words, Chisholm concedes the need for independent information to justify their application.

Chisholm, however, claims that words used noncomparatively may be applied without independent information. Words used noncomparatively may be homogenized to produce a single term applying to the state of a person. As Chisholm suggests, we might speak adverbially and say that a person believes that he is being appeared-to-redly or, equivalently, that he is sensing-redly. The hyphenated term is intended to characterize the subjective state of the person in question without implying that some thing is appearing to him and without implying any comparison of this state to any other. Thus, to say that one is sensing-redly does not entail that one is sens-

ing the way normal observers sense in daylight when they are seeing a red object. It may, of course, be true that one is sensing in that way, but it is not an analytic consequence of the term 'sensing-redly' used noncomparatively.

There is an argument to show that there must be a noncomparative use of words. To be able to describe a state in comparative terms, thus comparing it to other states, one must first be able to tell what the state is like in itself. Unless one can tell what this state is like in itself, one will be unable to compare it successfully to anything else. To be able to tell whether *A* is similar or dissimilar to *B,* one must first know what *A* is like, or any such comparison will be unreliable. Therefore, there must be a way in which one can tell what something is like, for example, a state of oneself, without comparing it to anything else. The noncomparative use of words enables one to record these noncomparative determinations.[7]

We agree with Chisholm that there is such a thing as a noncomparative use of words and ask whether this will sustain the thesis that beliefs expressed in noncomparative terms are justified without the need for independent information. For a person to be justified in believing that she is sensing in a certain way, that she is sensing-redly for example, she must have the information necessary to distinguish this manner of appearing from others. Perhaps, as Chisholm contends, the belief that one is appeared-to-redly does not entail any comparison of one's present state to any other. Nevertheless, it does entail that one's state is of a certain kind, and, to be justified in believing it to be of that kind, one must have the information needed to enable one to distinguish such a state from one of another kind. So we arrive at the conclusion reached earlier: to be justified in believing anything about a state or an object or whatnot, one always requires independent information.[8]

Semantics and Justification

The preceding reflections might seem to doom the foundation theory to epistemic oblivion. We have seen that the stockpile of infallible beliefs is epistemically inadequate to provide a justificatory foundation. We have now considered the possibility that fallible beliefs might be justified in themselves without the need for any independent information, and we have found this proposal wanting. The latter proposal may, however, be defended with some modification. The effective modification is to claim that some beliefs are justified in themselves just because of their semantics. The independent information required for such beliefs to be justified, it may be argued, is not *evidential* information used to justify the beliefs but *semantic* information required to understand the content of them.

Semantic information is information a person requires to understand the meaning of a word. A person who learns the meaning of a word, the word

'dog' for example, must acquire some information about what dogs are like in order to understand the meaning of the word. Such information required for an understanding of the meaning or content of a word is semantic information. If the required information is nothing more than semantic information needed to understand the meaning of a word and, hence, the content of the belief, the need for such information does not preclude the possibility of some beliefs being justified in themselves.

Meaning and Semantic Foundationalism

Again, we are faced with a solution to an epistemological problem based on semantics. We might call the position *semantic foundationalism*. Is the solution effective? To say that it is true by virtue of semantics that beliefs with a certain kind of content are justified in themselves is equivalent to saying that the epistemic principles according to which beliefs of that kind are self-justified are semantic principles or meaning postulates.[9]

Typical meaning postulates are principles such as the following:

All bachelors are unmarried.
All triangles have three sides.
All red things are colored things.

Such principles are true by virtue of the meaning of the words contained therein. Thus, the claim that certain beliefs are justified as a result of meaning postulates is equivalent to saying that an understanding of the content of the belief is sufficient to know that the beliefs are justified in themselves, just as an understanding of something being red is sufficient to know that it is colored. This is a consequence of semantic foundationalism.

Meaning and Skepticism

Is the thesis true of empirical beliefs? Are there beliefs about the empirical world, either the world of sensory experience within us or the world of perceptual objects outside of us such that an understanding of the content of the beliefs, of the meaning of the words used to describe them, is sufficient to know that the beliefs are justified in themselves? Or, do such beliefs require the confirmation of empirical evidence and, hence of auxiliary information, for their justification? How might one reply to the semantic foundationalist who claims that it follows from the content of beliefs and meaning of words that some such empirical beliefs are basic self-justified beliefs?

A skeptic might claim that no beliefs are justified or, more modestly, that the alleged basic beliefs are not justified. She may share most of our beliefs, but she does not share our epistemic convictions concerning what is evident, certain, justified, and so forth. If our fundamental question about

whether semantic foundationalism is correct is to be answered in the affirmative, then appeal to semantics—to the meaning of words—must suffice to refute the skeptic. Let us consider whether the meanings of words are sufficient to untie the skeptical knot.

If a person should claim that something is red but not colored, we might conclude that he was contradicting himself or else using the words in question in some peculiar way to mean something different from what is ordinarily meant. Should we conclude that the skeptic is using the word 'justified' in some peculiar way to mean something different from the rest of us when she claims that simple perceptual beliefs or beliefs about how things appear are not justified beliefs? Must the skeptic mean something different by epistemic terms simply because she speaks with a skeptical tongue? The answer to this question, happily for the skeptic, is that she need not mean anything different by these terms from the rest of us. The systematic difference in what she says from what the rest of us say may suggest that she means something different, but the conclusion cannot be forced on her.

We noted in the last chapter that there are two ways to account for the fact that others regularly say different things from what we say in the same situations. One way is to suppose that the words they utter mean something quite different when uttered by them. The other is to suppose that the words mean nothing different, but they differ from us in what they believe. When the skeptic utters epistemic words, we may either suppose that she means something different from the rest of us or we may suppose that her beliefs differ from ours. How can we show which supposition is true?

The situation is like the one in the previous chapter. Suppose the skeptic provides us with an explanation of why she speaks the way she does in terms of her beliefs about justification. For example, suppose she holds the view that only beliefs whose truth is guaranteed are justified and denies that the beliefs in question are ones whose truth is guaranteed. Her conclusion that such beliefs are not justified is most simply explained by these background beliefs and does not permit us to attribute semantic deviance to her remarks.

There may be no evidence of semantic deviance based, for example, on the way the skeptic understands the semantic relations between epistemic terms, such as 'know,' 'certain,' 'evident,' and 'justified,' composing what John Lyons and Adrienne Lehrer consider a semantic field.[10] Agreement about semantic relations between terms in a semantic field is a mark, if not a proof, of agreement in the meaning of terms. Moreover, there is nothing contradictory in our skeptic's skeptical claims. Nothing she says contradicts itself or anything else she says. The skeptic is consistently odd. The simplest explanation for the oddity of her remarks is that she believes that no one is justified in believing what they do unless they have some guarantee of the

truth of the belief. That is what she says and that is what she means. We should take her at her word.

Self-Justification and Necessary Truth

We have said that the skeptic is not contradicting herself, not affirming anything necessarily false, when she denies that alleged basic beliefs are justified. Can we give some argument to refute the claim of a foundation theorist who affirms, to the contrary, that it is true by virtue of the meaning of words that certain beliefs are self-justified, and it is contradictory to deny this? There is little credibility in the idea that when the skeptic says that a belief is not justified she is thereby saying something contradictory. Compare the skeptic's claim with the claim that there is a red object that is not colored. The latter, if taken literally, is impossible. There is no logically possible way in which an object can be red but not colored. The suggestion is logically incoherent. By contrast, the claim that a belief, any belief, is not justified, however contrary to our convictions, is logically coherent. We can understand how it is possible, for example, if there are no justified beliefs. The skeptic does not contradict herself. It would be plainly dogmatic and unwarranted to pin the label of inconsistency on her pronouncements.

Justification and Necessity

We can, moreover, go beyond general skeptical argumentation to refute the claim that the meaning of words sustains self-justification. At the same time, we can refute the related claim that principles of self-justification, those telling us that some kinds of beliefs are self-justified (simple perceptual beliefs for example) are justified in themselves. We noted above that the justification of alleged basic beliefs depends on possessing independent information on a given occasion that enables one to discern the difference between one state or thing and others on that occasion. Suppose that a person believes that she sees a red thing but lacks the information on that occasion enabling her to tell the difference between a red thing and a green one. Then she is not justified in believing that sees a red thing on that occasion because she lacks the information then that would enable her to distinguish red things from things of other colors.

This point must be formulated with some care, for the semantic foundationalist may plausibly argue that some competence to distinguish red things from things of other colors is required to understand the meaning of the word 'red' or the content of the belief that something is red. Suppose, however, that a person does have the general competence to distinguish red things from things of other colors but, on a given occasion, lacks the information needed to exercise the competence. If the person is using the word 'red' to describe some external quality of an object, the person might lack

information about whether the lighting is normal or abnormal, perhaps because he or she has entered a house of visual illusion. If the person is using word 'red' to describe a way of sensing, that is, noncomparatively, the person might lack internal information about whether that is the way of sensing described by that word perhaps because they are in a mentally confused state.

Semantic Foundationalism: A Counterexample

Imagine that a person, Mrs. Doubtword, understands the meaning of a word used to describe a way of sensing or appearing, the word 'red' as so used, for example. She has the general competence to apply the word correctly. She believes that the word applies presently, that is, that the way she is appeared to is red, but doubts her competence in this case because of confusion generated by a medication that she takes which sometimes creates occasional brief periods of linguistic and conceptual confusion. During such periods she unknowingly misapplies words and concepts rather like a slip of the tongue in speech. We might imagine that Mrs. Doubtword is perfectly competent at present, but she has taken medicine and is now feeling some side effects. These side effects are actually irrelevant to her linguistic competence, but, nonetheless, she has doubts about her linguistic competence at the present moment because she is experiencing these side effects of the drug. As a result, she fails to be justified in believing that 'red' describes the way she is being appeared to at this moment, though the word does describe the way she is appeared to and she is, in fact, competent at the moment. Mrs. Doubtword fails to be justified in believing the word applies because she recognizes that this might be one of those periods of linguistic and conceptual confusion to which she has been subject on past occasions when she felt the side effects of the drug.

Mrs. Doubtword's doubts about her competence are like the doubts of the person in the house of illusions when he is not deceived. Both have reason to doubt their competence because of the abnormality of a condition associated with the competence. In the case of Mrs. Doubtword the condition is internal, whereas in the case of the man in a house of illusions it is an external circumstance. Both may believe what their senses tell them, but both have doubts about their competence that undermine their justification for what they believe. Both lack the independent information that they are competent in the necessary way for being justified in believing what they do.

We can easily generalize this consideration. Put formally, suppose someone claims that for any *S*,

(*a*) *S* believes that *x* is *F*

necessarily or semantically implies, that is, entails

(*b*) *S* is justified in believing that *x* is *F*.

There is a recipe for constructing counterexamples that even a novice can follow. Suppose that

(*c*) *S* lacks the information on this occasion about whether he or she is competent to tell whether something is an *F* or not and doubts that he or she is competent to tell the difference on this occasion.

The conjunction of (*a*) and (*c*) obviously does not entail (*b*). A person who believes that *x* is *F* but lacks the information on this occasion about whether he or she is competent to tell whether something is an *F* or not and, moreover, doubts that he or she is competent to tell the difference, is obviously not justified in believing that *x* is *F*. If, however, the conjunction of (*a*) and (*c*) does not entail that (*b*), then (*a*) alone does not entail that (*b*). This is a consequence of the general principle that if a premise entails a conclusion, then the premise conjoined to anything else also entails the conclusion.

The argument illustrates a consideration introduced in the last chapter concerning how the opacity of some feature can undermine justification. We noted in that chapter that the opacity of infallibility leads to the result that infallibility does not yield justification when infallibility is opaque to the subject. Here we note that competence, though it falls short of infallibility, may also be opaque to the subject with the same result, namely, that competence does not yield justification. The reason in both cases is that the feature, infallibility or competence, may be concealed from the subject and fail to reveal the evidence required to justify the person in what he or she believes. Opacity extinguishes the light of justification.

Necessity and Prima Facie Justification

There are two foundationalist replies to this argument. One is that the justification semantically implied by belief is a prima facie justification only. A second is to abandon the semantic strategy and admit that justification is not semantically implied but only contingently implied. Let us consider the first line of reply. What does it mean? It means that it is semantically or necessarily true that if a person believes that *p*, then the person is justified in believing that *p* unless the justification is undermined. The class of beliefs of this sort would most naturally be specified in terms of the content of what is believed. Beliefs about one's present thoughts and feelings, or beliefs about some object or quality one sees directly before one or remembers having seen, are plausible candidates.

There are two objections to this line of thought. The first is already familiar. A skeptic who claims that no belief is ever justified, and hence that the alleged basic beliefs are not prima facie justified, appears to be perfectly consistent. If, however, it is consistent to deny that alleged basic beliefs are prima facie justified, then it is not true by virtue of the meaning of words or necessarily true that the beliefs are prima facie justified. The skeptic surely is consistent in what she claims, and therefore it is not true by virtue of the meaning of words or necessarily true that the beliefs are prima facie justified.

Moreover, the claim that certain kinds of beliefs are prima facie justified may lead to dogmatism concerning what beliefs are justified unless some reason, some independent information, is offered to explain why those beliefs and not others are prima facie justified. Suppose, for example, that I, being a partisan of the theory of paranormal phenomena, claim that if one person believes he has communicated with another person telepathically, then the person is prima facie justified in this belief. You might retort that people have such beliefs all the time, and that they are more often erroneous than correct. You might add, moreover, that people lack the information to tell whether or not they have communicated with another telepathically and, as a result, are incompetent to offer independent information to explain why those beliefs concerning telepathic communication are prima facie justified when other beliefs are not.

Suppose I reply that such considerations merely override the prima facie justification of such telepathic beliefs but fail to show that the beliefs are not prima facie justified. How could you reply? You will find it difficult, and the difficulty exposes the emptiness of the claim. Any argument to show that a belief is not prima facie justified may be interpreted as merely showing how the prima facie justification of the belief can be overridden. Any claim of prima facie justification can be rendered untestable and invulnerable to criticism but, alas, only at the cost of rendering it vacuous. If all reasons for denying that a belief is justified are construed as merely showing that the prima facie justification is overridden, then any claim that any belief is prima facie justified is rendered trivial. In this way, all beliefs, as well as beliefs of their denials, may be trivially affirmed to be prima facie justified. No epistemology flows through an empty pipe.

The principal difficulty for the foundationalist who claims that beliefs are prima facie justified is not that such claims are vacuous, though they may be vacuous, but that the claim that such beliefs are prima facie justified requires the support of an auxiliary assumption—of independent information—to explain *why* those beliefs are so justified and others are not. But the appeal to such information, which would, presumably, claim that such beliefs are, though fallible, nevertheless, a good guide to truth, good enough, anyway, to provide us with a good chance of avoiding error, re-

futes the claim that the justification of the beliefs is basic. The foundationalist claiming prima facie justification for some beliefs faces a dilemma. Either the claim that this particular belief is justified is supported by independent information indicating that the beliefs are a good guide to truth, or it is not. If it is so supported, then the justification of the beliefs becomes nonbasic because it depends on independent information. If it is not so supported, then the claim becomes dogmatic and vacuous. Either way, the attempt to sustain the justification of basic beliefs by appeal to prima facie justification fails.

Once again, it is important to notice the impotence of opaque features to provide justification, even prima facie justification, for basic beliefs. For it may well be that the beliefs that the foundationalist claims are basic and prima facie justified are, in fact, a good guide to truth. But if the fact that they are a good guide to truth is opaque to the subject, then the fact is impotent to provide the subject with justification unless the subject possesses the information that the beliefs are a good guide to truth. The need for that information shows that the justification for the beliefs is not basic, not inherent in the belief, but dependent on auxiliary information.

In summary, the claim that alleged basic beliefs are necessarily prima facie justified is either vacuous or false. If the claim is not vacuous, then the possibility of a person lacking the information enabling him or her to tell whether the alleged basic belief is a good guide to truth, suffices to refute the contention that it is necessarily true that the belief is prima facie justified. Cognitive opacity based on ignorance of required auxiliary information does not override prima facie justification but excludes justification in the first instance.

Contingent Self-Justification

At this point, the fallible foundation theorist may retreat from semantics to the claim that the self-justified beliefs are self-justified as a contingent matter of fact. There is something plausible about the proposal that some beliefs are justified in themselves, that they are, so to speak, the first premises of inference. There are some beliefs that are justified without being consciously inferred from any other. When I see a red object, I believe I see something red without consciously inferring that from anything else. When, however, we ask why that belief is justified, the answer reveals a dependence on independent information, as noted above, and shows that the belief, though justified without conscious inference from anything else, is not self-justified in the sense required by the foundation theory. The justification of the belief depends on other information, to wit, information that I am competent to discern truth in such matters and, consequently, that belief in this case is a good guide to truth.

These considerations might lead us to doubt the tenability of even a modest fallible foundationalism, affirming that, as a matter of contingent fact, some beliefs are justified in themselves independently of other information and belief. It appears that the justification for accepting anything we believe depends on other information—general information that enables us to obtain truth and avoid error. William Alston and James Van Cleve have argued, however, that self-justified basic beliefs may provide the basis for justifying general beliefs which, in turn, may be used to explain why the basic beliefs are self-justified.[11] We infer that the basic beliefs are almost always true and thereby obtain the means to explain why we are justified in accepting the basic beliefs in the interests of obtaining truth and avoiding error.

The question for the foundation theorist, however, is whether the justification of our basic beliefs depends, at the time at which they originally arise, on our beliefs about our competence to tell what is true and what is not in such matters, or upon the information that we have such competence. I believe that I see something red. Someone alleges the belief to be basic. It may well be that I have not consciously inferred this belief from any other, but the justification for accepting the belief seems to depend on the assumption, on my assumption, that I can tell a red thing when I see one. If I have no idea whether I can tell a red thing when I see one, then, even if I can, my belief that I see something red lacks the sort of justification required for knowledge.

To become convinced of this, suppose that I am asked to try to distinguish real diamonds from fakes by looking at them through a magnifier. Imagine, though I have no idea that it is so, that all my identifications are correct. Real diamonds look different to me from zircons and other imitations, and so, when I believe I see a diamond with the magnifier, I do indeed see a diamond. I have never had my judgments checked, however, and I really have no idea that I always see a diamond with the magnifier when I believe I do or that I have the competence to tell a real diamond from a fake. I am not justified in accepting my belief that I see a real diamond through the magnifier, even if I do, because I have no evidence that any of my identifications are correct.

If, on the other hand, I have evidence, from checking my identifications against an authoritative list, that I have the special competence to tell a real diamond from a fake in terms of some noted difference in their appearance when viewed though a magnifier, I might justifiably accept that I have the competence to identify a real diamond in this way. On that assumption, when I next identify something as a diamond in this way, I am justified in believing that I see a diamond. Once I have justified the belief about my competence in identifying diamonds, I may come to believe that I see a diamond through the magnifier without inferring this from the information about my competence. I may simply form the habit of believing that I see a

diamond when I identify something as a diamond with the magnifier. The justification for my belief that I see a diamond depends, nevertheless, on my background information that I have the special competence to identify a diamond with a magnifier when I see one. Similarly, when I believe that I see something red, the justification I have for accepting that belief in the interests of obtaining truth and avoiding error depends on my justified background belief that I have the competence to tell a red thing when I see one, though I do not *infer* that I see a red thing from that background belief. I may, at first, infer that I see a diamond or a red thing from a premise about my competence in diamond or red-thing identification but, subsequently, a habit of belief replaces inference. Nevertheless, my justification for accepting these beliefs depends on background information about my competence to obtain truth and avoid error in my doxastic commitments.

The moral of the story about basic beliefs was suggested at the beginning of our examination of the foundation theory. If alleged basic beliefs do not guarantee their truth, then the justification for accepting those beliefs in the quest for truth must depend on other information or beliefs. The alleged basic beliefs will fail to be self-justified. All justified acceptance of beliefs, even the small number of infallible beliefs, will depend on some background information we possess about our competence to determine whether the contents of the beliefs are true or false. Our knowledge depends on this background information.

Probability and Justification: Fallibilistic Foundationalism

We also noted at the outset that if our beliefs are fallible and do not guarantee their own truth, then such beliefs entail some risk of error. If such beliefs carry a risk of error, then, assuming we seek truth and eschew error, we must be assured that the risk is worth taking. The risk of error is the probability of error. The fallible foundationalist allows that there is a risk of error, but this risk, or *probability,* of a belief being erroneous must not be too great or we shall not be justified in accepting the belief for the purposes of obtaining truth and avoiding error. Thus, any belief we are justified in accepting, whether basic or nonbasic, must have a sufficiently high degree of probability to justify acceptance.

The preceding formulation of the problem takes us to the most fundamental issue with respect to fallible foundationalism, one concerning the probability of truth. We suppose the fallible foundation theorist to have given up the attempt to find any guarantee of truth in the justification of basic or nonbasic beliefs. For the foundation theory to succeed, the basic beliefs of the theory must be highly probable and must render nonbasic beliefs highly probable as well. If the evidence formulated in self-justified ba-

sic beliefs is to justify nonbasic beliefs, then the truth of nonbasic beliefs must be highly probable on the evidence of the basic ones. Nonbasic beliefs are, therefore, justified in the requisite sense only if they are at least highly probable on the evidence of basic beliefs.[12]

Three Concepts of Probability

Philosophers and logicians have distinguished a number of different conceptions of probability.[13] Among these are the frequency concept, the logical concept, and the subjective concept. We shall consider them all and ask what application each concept has to the problem of the justification of nonbasic beliefs. First, we consider the frequency concept. A frequency probability statement expresses the numerical frequency with which members of class *A* are also members of class *B,* for example, the frequency with which smokers are people who suffer heart disease. What exact interpretation is given to such statements is controversial. Some philosophers have interpreted such statements as expressing the limit of the relative frequency in an infinite series. Other philosophers have interpreted the frequency statement as expressing a propensity of members of one class to be members of a second class.

However, the crucial feature of such statements for the purposes of our discussion is that all such statements turn out to be very general contingent statements about the world. Consequently, if such probability statements enter into the justification of our beliefs, whether basic or nonbasic, the question arises of whether we must be justified in accepting the frequency statements and, if so, how we can be justified in accepting them in terms of fallible foundationalism. Are they basic beliefs? If so, how can they be self-justified? If they are not basic, what basic beliefs justify them?

Frequencies and Justification: An Objection

To appreciate the problems associated with frequency probabilities and the justification of nonbasic beliefs on the evidence of basic beliefs, let us return to the special case of justifying statements about external objects on the basis of sense-data statements, for example, the belief about there being a tomato in front of me on the evidence of the reddish, roundish, bulgy sense data. We noted that this evidence by itself does not logically guarantee the truth of the claim about the tomato, but a fallible foundationalist might argue that the evidence renders the claim about the tomato highly probable in a frequency sense of probability. The frequency with which tomatoes accompany sense data of this sort is very high.

There is a classical objection to the strategy of employing frequency probabilities to justify statements about external objects on the evidence of sense-data statements. It is that, in order to determine the truth of the prob-

ability statement concerning the frequency with which some special variety of sense data are accompanied by the existence of external objects of a specified kind, we would already have to have determined the frequency with which the special variety of sense data were accompanied by the specified kind of external objects in a fair sample of observed cases. By analogy, to determine that smokers get heart disease 30 percent of the time, one would have had to determine the frequency with which smokers get heart disease in a fair sample of cases.

To know the frequency of the presence of external objects (tomatoes, for example) in the sense-data sample, however, one would have to know what the frequency probability statement was supposed to enable us to know, namely, that beliefs about the external objects, the tomatoes, are true. The attempt to justify statements about external objects by appeal to frequency statements is, therefore, futile. It presupposes that we already know and are already justified in believing exactly those statements—the ones concerning external objects—that the frequency probabilities were supposed to enable us to justify. That is the classical argument.

It is a plausible argument, indeed. If we grant that the only method we have of knowing that a frequency statement is true is by inference from what we know in a sample, and that the frequency statement must be known for the truth of it to sustain the justification of the nonbasic beliefs, then the argument is decisive. It is most natural to make just this assumption, for how else are we to know that the frequency statement is true except by appeal to a sample of the relevant sort? And if we do not know that it is true, how can it justify any belief? How can a statement, even a true one, of which we are ignorant, justify us in believing anything?

Truth Frequency and Externalism

There are two distinct lines of reply. A fallible foundationalist influenced by externalist theories might reply that frequencies suffice to yield justification and knowledge even when we are totally ignorant of them. In terms of our example, the mere high frequency with which reddish, roundish, bulgy sense data correlate to the existence of tomatoes suffices to justify our belief in the existence of the external object, the tomato, on the evidence of the sense data—even when we are ignorant of the frequency. It is, on this view, the frequency rather than our knowledge or justified acceptance of it that enables our basic beliefs to yield justified nonbasic ones. Indeed, such a foundationalist enthralled with the advantages of externalism might even go on to claim that the justification of basic beliefs is the result of their truth frequency rather than of any information we have about such frequencies. Truth frequency itself greases the slide of justification.

A detailed account of externalism will occur in a later chapter, but it is easy enough to see why traditional foundationalists have not been seduced

by the ease of this solution of their problem. We have already considered the critical rejoinder in the example of the person who, without knowing it, succeeds in correctly identifying diamonds with a magnifier. The person might notice some special appearance of some of the stones she examines, a special luster, for example, and believe that the stones with this appearance are the diamonds. If she happens, as luck would have it, to be correct, then the truth frequency of her beliefs about which stones are diamonds will be high, though she will be ignorant of this fact. But since she is ignorant of the fact that the appearance she notices is the appearance of the diamond, ignorant, that is, that this is the way diamonds appear, the truth frequency leaves her without justification for believing that the stones that appear in the way she notices are diamonds. Her beliefs about which stones are diamonds are correct, but she is not justified in accepting them, nor does she know that they are true. Ignorance of the frequencies is the sticking spot on the slide.

A foundationalist, who acknowledges that frequencies of which we are totally ignorant yield neither justification nor knowledge, might reply instead that at least some such frequency statements are basic, ones we are justified in believing without being justified by any other statements. A philosopher willing to defend the foundation theory by arguing that the frequency statements needed for basic statements to justify nonbasic ones are themselves basic statements may avoid the need to justify the frequency statements by appeal to anything else. Being self-justified, the beliefs about frequency probabilities can bridge the justificatory gap between basic beliefs about sense data and nonbasic beliefs about external objects.

A skeptic might claim that such a foundation theory begs the question against skepticism. Empiricists might protest that if the frequency statements are assumed to be basic, then such a foundation theory abandons empiricism by supposing that general statements, the frequency statements, are justified without the confirmation of particular observations. These objections have no force against a foundation theorist who is willing to accept the inadequacy of skepticism and empiricism, however. There is no internal inconsistency in the attempt to employ frequency probabilities to explain how nonbasic statements of a foundation theory are justified by basic beliefs, provided that we are willing to adopt beliefs about frequencies as basic. Is it tenable to solve the problems of foundationalism in this way?

Unfortunately not. The defect in the proposed solution becomes apparent when we consider the question, Why are we justified in accepting basic beliefs about the frequencies? When the beliefs about the frequencies are held to be basic beliefs by the fallible foundationalist, the only answer is that such beliefs are self-justified. This answer has the twin disadvantages of being uninformative and incorrect. We posed the question to find out

why we are justified in accepting beliefs about the frequencies. The answer—that the beliefs are basic—amounts to saying that we just are justified and fails to inform us as to why we are. Since it is not at all obvious why such beliefs should be justified in themselves, the answer is uninformative. We have argued above, moreover, that uninformative answers to questions about why we are justified in our beliefs fail to reveal the evidence needed to support the justification and confront us with the problem of cognitive opacity.

The answer that beliefs in such frequencies are basic, in addition to being uninformative, also is obviously incorrect. We are convinced that there is a better answer, namely, that we have evidence from past experience of the frequency of truth. The reason that we are justified in accepting that we see tomatoes, chairs, doors, hands, and other familiar perceptual objects on the evidence of how they appear to us is that such appearances have rarely led us to error. We are justified in accepting such perceptual beliefs on the basis of appearances because such beliefs are almost always correct and almost never in error. That is why cautious perceptual beliefs about seeing tomatoes on a table directly in front of us are justified, whereas more speculative ones about seeing water on a highway far away are not. The appearances of the former almost always produce true beliefs about the presence of tomatoes, whereas the appearances of the latter frequently produce false beliefs. The claim that beliefs about such frequencies are basic is in error. They are justified, when they are, by the evidence of our past successes and failures.

Logical Probabilities and Justification

It is, therefore, worthwhile to ask whether some other conception of probability, the logical or subjective conception, is better designed to supply a bridge of probability between basic and nonbasic beliefs. In fact, both the logical and subjective conception at first appear better adapted to this purpose than the frequency conception. We cannot undertake a detailed examination of either, but a rough description will suffice to convey the idea of each and the similarity between them.

The logical probabilities are formed by considering the basic logical alternatives and assigning probabilities to them. A standard example would be that of a normal six-sided die. From a simple description of the die and prior to having observed it being thrown, one would assign a prior or antecedent probability of 1/6 to a face turning up on a normal throw of the die. Why? Logic alone tells us that there are six sides and, in the absence of any empirical information about how the die behaves when thrown, we regard them as equally likely to turn up. You might be inclined to think that some empirical information about the behavior of our die is influencing this assignment of probability. Suppose, on the other hand, that you were igno-

rant of such empirical information. With only logic to guide you, would you not assign the probabilities in the manner indicated?

That is the idea behind the logical conception, which may, moreover, be applied on a global scale. Consider the basic, logically alternative ways one might describe the world and assign probabilities to these alternatives. From that assignment, one may define a conception of probability, a logical conception sufficient to define all probabilities. That is the theory; all you need is logic and definitions. The probabilities are all consequences of the definitions.

You might, of course, object that all the probabilities will depend on how we assign them to the basic alternatives. It might seem obvious enough that we should assign equal probabilities to various faces of the die turning up when the die falls on one of the faces, but must we assign equal probabilities to the various total descriptions we might give of the universe? And might there not be different ways of providing total descriptions resulting in different ways of assigning probabilities? In fact, logicians have shown that there are infinitely many different ways of assigning logical probabilities.[14] This diversity provides a motive for the subjective theory.

Subjective Probabilities and Gambles

One is left by the logical theory with many possible ways of assigning the initial or prior probabilities that are all consistent with the basic postulates of probability theory, and the choice between them appears subjective. The subjectivist simply embraces the result. We may assign prior probabilities as we choose as long as the assignment is consistent with our own behavior and our behavior is rational. What sort of behavior is connected with the assignment of probabilities? Our choices and, most importantly, our betting choices.

Consider a bet in a game of chance in which a person must pay $.75 for a gamble, receiving $1.00 for a win and nothing for a loss. Imagine you are offered the chance to take the bet, that is, pay $.75 for the gamble, receiving $1.00 if you win or nothing if you lose, or give the bet to someone else, that is, receive $.75 and pay the other $1.00 for a win or nothing for a loss. (The gamble might be a draw from an ordinary deck of cards with the stipulation that you lose if a heart is drawn but win otherwise.) Suppose you find that you are indifferent to taking the bet or giving it. In that case, your behavior is rational just in case you assign a probability of .75 to winning the gamble. If you assign a higher probability, you should rationally choose to take the bet, and if you assign a lower probability, you should rationally choose to give the bet. If you are indifferent, you should rationally consider the bet fair, that is, the probability of winning to be exactly .75.

In general, the subjectivist allows any assignment of probabilities that is consistent with rational choice. It turns out, interestingly enough, that your

assignment of probabilities will be consistent with the postulates of the theory of probability if and only if using your probabilities for betting would not allow someone betting with you and knowing your probabilities to bet in such a way as to be certain of winning money from you regardless of the outcome of the gambles. If someone could be certain of winning money from you by betting with you on the basis of the probabilities you used to determine fair bets without knowing the outcome of the gambles, then your probability assignment is incoherent and irrational.

One simple way for this to occur would be if you assigned two different probabilities to the same thing, for example, .75 and .65, to winning the gamble considered above. In that case, someone betting with you could give you \$.65 for the gamble, receiving \$1.00 if she wins and nothing if she loses, since that is fair if the probability of winning is .65, as you said it was, and, at the same time, demand \$.75 of you for the gamble, giving you \$1.00 if you win and nothing if you lose, since that is fair if the probability of winning is .75, which you also said it was. Whatever the outcome of the gamble, whether a heart is drawn or not, she will have gained \$.10 and you will have lost \$.10. Your assignment of probabilities results in irrational behavior when they are used as a betting quotient. On the contrary, if they do result in rational behavior when they are used as a fair betting quotient, they are coherent and consistent with the theory of probability. That is all a subjectivist thinks it is reasonable to require of a probability assignment.

Logical and Subjective Probabilities: Advantages and Objections

What are the advantages of a logical or subjective conception of probability? You can ascertain logical or subjective probabilities by mere reflection without determining any frequencies. Thus, the information concerning such probabilities does not lead directly to the impasse we noticed with respect to information concerning frequency probability statements; to wit, that the employment of such information to justify nonbasic beliefs on the evidence of basic beliefs presupposes that we already know and are justified in accepting nonbasic beliefs to be true. Logical and subjective probability statements can be known without knowing anything about frequencies because they do not tell us anything about frequencies. It might thus seem that a fallible foundationalist should avail himself of these nonfrequency probabilities to span the justification gap between basic and nonbasic beliefs.

Nevertheless, the encouragement offered by these probabilities has a hidden defect. Since there is an infinite number of ways in which we might assign such probabilities, all of which are coherent and consistent with the theory of probability, we face the problem of establishing a connection be-

tween probability and truth. There may, after all, be none. Granting that our nonbasic beliefs are highly probable on the evidence of our basic beliefs, a detractor may query, How can your assigning a high probability to a nonbasic belief on the evidence of a basic belief justify us in accepting the nonbasic belief as true when there are many other ways of assigning probabilities under which the same nonbasic belief would have a low probability on the evidence of the same basic beliefs?

One reply to such a query is that the logical or subjective probability assigned is our *estimate* of truth frequency, but this reply, as it stands, is inadequate. The question remains whether it is a good and reasonable estimate. The answer must be that it is a basic, self-justified belief that the probabilities in question are reasonable estimates of truth frequencies because only then will the probabilities become appropriate to the requirement of justification that acceptance should provide us with a good chance of obtaining truth and avoiding error. Here we encounter a strategy similar to the one employed in the defense of the use of frequency probability statements in the justification of nonbasic beliefs by basic ones. We close a gap in the justification by filling it with a basic belief concerning frequencies.

We are back to where we stood when we considered the frequency conception of probability. There the problem was to explain how we are to ascertain that such frequency probability statements are true, for, if they are true, they establish a frequency relation between the truth of basic and nonbasic beliefs. In the case of logical and subjective probabilities, the problem of ascertaining probabilities can be solved by reflection and calculation alone. Though the problem of ascertaining probabilities is less, we are nevertheless confronted with the problem of relevance. That is to say, here we must show that probability is relevant to the truth of the nonbasic statements, and, more specifically, to the frequency with which nonbasic beliefs of a specified kind are true when basic beliefs of a certain sort are true.

If we cannot show that there is a correlation between the high logical or subjective probability of a nonbasic belief on the evidence of a basic one and the truth of the nonbasic belief on the evidence of the truth of the basic belief, then such high probability is irrelevant for the purpose of our justifying nonbasic beliefs on the evidence of basic ones. Thus, in the case of both logical and subjective probability, we need additional information about the correlation between these probabilities and truth frequencies to render them relevant to justification aimed at obtaining truth and avoiding error. Our information about the correlation presupposes information about the frequencies which, for the reasons we noted when considering the frequency conception of probability, must be assumed to be basic. We are, therefore, no better off with logical or subjective probabilities than with frequency probabilities. The problem is the same: the foundationalist must postulate that beliefs about the frequencies are basic, that is, self-

justified, and such postulation, as we noted, seems as incorrect as it is uninformative.

Probability, Truth, and Basic Belief

The conclusion is that whatever conception of probability we adopt, to render the probability relevant to justification of the required sort, we need the assumption that probability is a guide to truth. If we are justified in accepting such an assumption on a foundationalist theory, it must be because it is a basic belief. The underlying problem is like the one we uncovered in considering the justification of basic beliefs, to wit, that the justification of nonbasic beliefs, like the justification of basic beliefs, depends on some general information that tells us when our particular beliefs are a trustworthy guide to truth and when they are not.

In the case of basic beliefs, we require independent general information about when our basic beliefs are a trustworthy guide to their own truth. In the case of nonbasic beliefs, we require independent general information about when basic beliefs are a trustworthy guide to the truth of nonbasic beliefs. In both cases, the foundationalist must regard the independent general information as basic belief. This device raises serious problems, as we have noted.

First of all, how are we to avoid the charge of being arbitrary in what is claimed to be basic? The foundation theory was to provide a safe structure of justification where everything was based on a foundation of self-justified basic beliefs. Now we see that in order to construct this edifice—indeed, in order to lay the first foundation stone, as well as to lay the second upon the first—we need to assume as basic a justificatory superstructure of general information concerning the truth frequency of basic and nonbasic beliefs. The foundation, instead of consisting of particular beliefs that run a minimal risk of error, consists of general beliefs that suffer all the hazards of error common to general conceptions. Such hazards may prove inescapable, but foundationalism cannot claim security from them as a special advantage offered by the theory.

Finally, as we have noted concerning frequencies, it is difficult to believe that such general beliefs are not justified by other particular beliefs. Within a foundation theory, these general beliefs cannot be justified by particular beliefs without arguing in a circle. But the particular beliefs cannot be justified by those general beliefs without arguing in a circle, either. This suggests, contrary to the foundation theory, that the justification of both kinds of statements may be reciprocal, that each justifies the other as a result of cohering with a system of beliefs containing particular beliefs about what we experience, as well as general beliefs about our competence to discern truth from error and the frequency of our success in so doing. To concede

this, however, is to give up the foundation theory and embrace the coherence theory instead. We shall turn to such a theory in subsequent chapters.

Summary: Competence, Success, and Coherence

The fallible foundationalist does not require that justification guarantee truth, but a person must accept that it is correlated with truth to obtain justification. Acceptance of the correlation is indispensable because a purpose of justification is the attainment of truth and the avoidance of error. To ensure the truth correlation for the justification of nonbasic beliefs, we found need of general information concerning the frequency correlation between the truth of basic beliefs and the truth of the nonbasic beliefs they justify. The beliefs concerning such a correlation will turn out to be basic on a foundation theory, but this proposal appears incorrect.

To ensure the truth correlation for the justification of basic beliefs, we found need for independent information about the competence of the person to discern truth from error in the subject matter of those beliefs. As a result, the alleged beliefs turned out not to be basic. They depended for their justification on independent information about the competence of the person to distinguish truth from error. To say that the person is competent to distinguish truth from error in some subject matter implies that the person will be successful with a high frequency in accepting what is true and avoiding accepting what is false in that subject matter. So, again in the case of basic beliefs, what turns out to be basic is not the particular beliefs about what we see and think and feel but a general belief about the frequency of success in believing what is true and avoiding believing what is false. The fallible foundationalist must assume justified acceptance of a general truth correlation, and the only way for him to obtain it is to postulate it as basic. Such postulation may provide us with an account of the justification of both basic and nonbasic beliefs and, therefore, satisfy a desire for parsimony. There is something more important in philosophy than parsimony, namely, understanding. It is here that such postulation of justified acceptance of the truth correlation as basic fails us.

Why are we justified in accepting that our beliefs about our thoughts, feelings, and simple perceptions are almost always true while other kinds of beliefs are more frequently in error? Why are we justified in accepting that some nonbasic beliefs are almost always true, given the evidence of basic beliefs, while other beliefs are most frequently in error? A foundationalist precludes himself from providing an informative answer that is transparent to the subject of the beliefs when he postulates that the beliefs about such frequencies are basic. The truth frequency of the beliefs will be opaque to

the subject of the belief and remain impotent to justify the belief for the subject. Postulating that the general belief about our competence, trustworthiness, or success is basic is a refusal to answer the questions about why we are competent, trustworthy, and successful in our quest for truth. Postulation has all the advantages over philosophical explanation that theft has over honest toil.

Our legitimate hunger for philosophical explanation remains unsatisfied. How can it be satisfied? With few exceptions, justified acceptance of the sort required for knowledge depends on information concerning our competence, trustworthiness, and, in short, success frequency in discerning truth and detecting error. What we are justified in accepting about such success frequency appears not to be a basic belief, however, contrary to what the fallible foundationalist must aver. Our justification for what we accept about our success frequency depends instead on our information about our successes and failures in the quest for truth. In this way, justified acceptance of any belief will depend on a background system of information about our competency and incompetency, about successes and failures, in our attempt to find truth and elude error in what we accept about the world and our relation to it. That is the thesis of the coherence theory to which we shall now turn.

Introduction to the Literature

Mark Pastin presents fallible foundationalism in "Modest Foundationalism and Self-Warrant," and similar theories are articulated and discussed by Ernest Sosa in "Epistemic Presupposition" and William Alston in "Two Types of Foundationalism." An account of how epistemic principles might be justified in a foundation theory is contained in James Van Cleve's important "Foundationalism, Epistemic Principles, and the Cartesian Circle." See also Richard Foley's important discussion of foundationalism in *The Theory of Epistemic Rationality.* The most distinguished defender of the foundation theory is Roderick Chisholm in his various books, including *Theory of Knowledge,* 3d ed. A good introduction to the subjective or personalist theory of probability is to be found in Richard Jeffrey, *The Logic of Decision.*

Notes

1. Thomas Reid, *Essays on the Active Powers of Man* from *The Works of Thomas Reid, D.D.,* ed. Sir William Hamilton (Edinburgh: James Thin, 1895), 617.

2. This argument is derived from Chisholm in *Perceiving: A Philosophical Study* (Ithaca, N.Y.: Cornell University Press, 1957), chap. 5, "Justification and Perception," 54–66.

3. Thomas Reid, *Essays on the Active Powers of Man.*

4. See, for example, Alvin Goldman, "What Is Justified Belief?" in *Justification and Knowledge,* ed. G. S. Pappas (Dordrecht: Reidel, 1979); and Goldman, *Epistemology and Cognition* (Cambridge: Harvard University Press, 1986), chap. 5.

5. Gilbert Harman, "How Belief Is Based on Inference?" *Journal of Philosophy* 61 (1964): 353–59.

6. R. M. Chisholm develops his doctrine concerning the noncomparative use of words in *Perceiving,* chap. 4, and he discusses it again in *Theory of Knowledge* (Englewood Cliffs, N.J.: Prentice-Hall, 1966), chap. 2.

7. Chisholm offers such an argument in *Theory of Knowledge,* chap. 2, p. 36 n. 20.

8. This argument has been influenced by a similar argument in W. F. Sellars, "Empiricism and the Philosophy of the Mind," in his book *Science, Perception, and Reality* (London: Routledge & Kegan Paul, 1963), sec. 18, 146–47.

9. See J. L. Pollock, *Knowledge and Justification* (Princeton: Princeton University Press, 1974).

10. See John Lyons, *Structural Semantics* (Oxford: Blackwell, 1963), chap. 4; Adrienne Lehrer, "Semantic Cuisine," *Journal of Linguistics* 5 (1969): 39–55; and Lehrer, *Semantic Fields and Lexical Structure* (Amsterdam: North-Holland, 1974).

11. James Van Cleve, "Foundationalism, Epistemic Principles, and the Cartesian Circle," *Philosophical Review* 88 (1979): 55–91; William Alston, "Two Types of Foundationalism," *Journal of Philosophy* 73 (1976): 165–85.

12. This thesis may be questioned by those who defend a nonprobabilistic account of inductive support, for example, L. Jonathan Cohen in *The Implications of Induction* (London: Methuen, 1970); and *The Probable and the Provable* (Oxford: Clarendon, 1977).

13. The frequency concept is articulated by Hans Reichenbach in *The Theory of Probability* (Berkeley: University of California Press, 1949); the logical concept is developed by Rudolf Carnap in *The Logical Foundations of Probability* (Chicago: University of Chicago Press, 1972); and the subjective concept is argued by Richard Jeffrey in *The Logic of Decision* (New York: McGraw-Hill, 1965).

14. See Jaakko Hintikka, "A Two-Dimensional Continuum of Inductive Methods," in *Aspects of Inductive Logic,* ed. J. Hintikka and P. Suppes (Amsterdam: North Holland, 1966), 113–32; and Rudolf Carnap, *The Continuum of Inductive Methods* (Chicago: University of Chicago Press, 1952).

5

The Explanatory Coherence Theory

WE SHALL NOW CONSIDER an alternative theory of justification, according to which justification is a systematic relation of *coherence* among beliefs belonging to a system. A coherence theory affirms that a belief is justified if and only if it coheres with a system of beliefs. The following is a schema for a coherence theory of justification:

> *S* is justified in accepting that *p* if and only if the belief that *p* coheres with other beliefs belonging to a system *C* of beliefs of kind *k*.

This schema raises one question immediately and a second is hardly concealed. The first question concerns the relation of coherence. What is coherence? In what way must a belief *cohere* with other beliefs belonging to a system of beliefs in order to be justified? The second is, What kind of system is kind *k*? That is, what sort of system of beliefs makes a belief justified when the belief coheres with others in that system? Definite answers to these questions are needed to convert our schema into a substantive theory.

The Regress, or the Circle

Before attempting to answer these questions, however, an objection to any sort of coherence theory must be considered. It has been argued that no coherence theory is feasible. The argument purports to demonstrate the unavoidability of basic beliefs, and hence of a foundation theory, for an adequate theory of justification. It affirms that unless some beliefs are basic, the justification of all beliefs must inevitably lead either to an infinite

regress or to a circular argument. Its conclusion is that either consequence is epistemically intolerable, and therefore we must uphold the foundation theory.

This argument needs more precise articulation, which runs as follows. If a person is justified in a belief on the basis of evidence, then appeal to that evidence would constitute a correct answer to the question, How do you know? Now suppose no beliefs are basic. Then every justified belief is so justified by appeal to evidence, but evidence must itself be justified belief and therefore must also be justified by appeal to evidence. This means that every justified belief must be justified by some other, thus leading either to an infinite regress, or to a justificatory circle. If both those alternatives are unacceptable, then there must be some basic beliefs.

This argument must be met or the project of constructing a coherence theory will be doomed from the outset. Fortunately for the coherence theorist, the argument is defective. Not all justified beliefs need to be justified by *appeal to evidence.* Appealing to evidence is an activity or process that occurs over time. Being justified in believing something is a state that exists at a time and need not result from the activity or process of appealing to evidence over time. A person may be justified in accepting something because of the way in which the information that she has coheres, or fits together to support what she accepts. Thus, for example, when my friend enters, I may have information about how my friend looks and about where I am that justifies me in accepting that my friend is now in Tucson without needing to explicitly appeal to the evidence. A belief such as this one may be justified for a person because of some relation of the belief to a system of information to which it belongs, the way it coheres with the system, just as a color in a painting may be beautiful because of some relation of the color to other components in the painting, or as a piece in a puzzle has place in it because of the way in which it fits with the other pieces.

It may yet be objected that if a belief is justified when no belief is self-justified, then a person must be able, at least in principle, to carry out the justification. That is, she must be able to justify the belief by appeal to evidence, to justify her belief in that evidence by appeal to other evidence, to justify her belief in that evidence by appeal to still other evidence, and so forth. One reply to this objection is that a person might in principle be able to carry out each step of this justification without being able to carry out the entire process. As an analogy, a person might be able to add one to each number without being able to carry out the whole process. It would be mistaken to infer that there is some number to which a person is unable to add one from the fact that she is actually unable to carry out the infinite task of adding one to each number. Similarly, it would be a mistake to conclude that a person is not justified in any belief from the fact that she is unable to

carry out the infinite task of justifying every belief to another. Hence, the regress argument fails.

Nevertheless, it is only fair to point out one further foundationalist objection to the foregoing. Some beliefs surely appear to be such that, though they are justified, one cannot justify them by appeal to evidence. For example, a person might justify her belief that she sees an apple by appeal to the evidence that there is an object before her that looks red and apple shaped, and she might justify her belief that the object before her looks this way by appeal to the evidence that she *thinks* that there is an object that looks this way; but, eventually, she must reach the point where no further evidence can be elicited. We would thus come to some justified belief that the person would be unable to justify by appeal to evidence. This objection takes us to the heart of the argument.

The reply to the objection is twofold. First, it must be noted that justification is ordinarily justification to someone else, and whether a justification given to someone suffices will depend on what that person is willing to grant. If he is willing to grant that the object in the distance is Argile Hill if it looks like the hill in a picture, one need only show him that the object looks like the one in the picture to justify one's belief that the object is Argile Hill. If, on the other hand, he doubts that the object in the picture is Argile Hill, then justification will have to be extended. Hence, there is a pragmatic element in justification depending on the epistemic situation of the person to whom the justification is directed.[1] Nevertheless, one might hold that, whatever the pragmatics of everyday justification of something by one person to another, there is a kind of justification aimed at obtaining truth and avoiding error that requires that only those things should be taken for granted that are already justified in terms of those objectives.

The second reply is, therefore, fundamental. Even when no evidence is offered to justify a belief, the belief may be justified by the relation of the belief to a person's system of beliefs. A person's system of beliefs may justify her belief because of the way in which the system of beliefs supports her belief without anyone appealing to the system of beliefs to justify the belief. For example, a person's system of beliefs may imply that if a person believes that something looks red to her; then, unless there is some unusual condition or circumstance, the explanation for why the person believes something looks red to her is that something *does* look red to her. People ordinarily believe that something looks red to them because something looks red to them! It might seem strange for a person to appeal to such a consideration as evidence that something looks red to her, but it might be such an explanatory relationship of her belief to her background system of beliefs that makes her justified in believing that something looks red to her, nonetheless. This does not make the belief self-justified, however, even

though it may be noninferential. The belief is not justified independently of relations to other beliefs. It is justified because of the way it coheres with other beliefs belonging to a system of beliefs.

The foregoing reflections show that a coherence theory is possible and a foundation theory unnecessary. We have yet to explain what coherence is or what sort of a system a belief must cohere with to become justified. We now turn to answer these questions.

The Traditional Answer: Coherence As Implication

Let us begin with the relation of coherence. Some defenders of the coherence theory conceived of the relationship of coherence as a relation of necessary connection.[2] Thus, a belief that *p* coheres with other beliefs of a system C if and only if *p* either necessarily implies or is necessarily implied by every other belief in C. Suppose, however, that we have a system of beliefs that is logically consistent, contains some logically contingent statements, and is such that every statement in the system either necessarily implies or is necessarily implied by every other statement. We can easily form another system having these same characteristics by taking the contingent statements in C, negating them, and forming a new system containing the negations of the contingent statements in C, together with whatever noncontingent statements may have been contained in C. This new system will be consistent if C was, and it will be such that every statement in the new system either necessarily implies or is necessarily implied by every other statement. The new system, though just as coherent as the old in terms of the necessary connections between statements, will tell us exactly the opposite about the world. Every contingent statement in one system is negated in the other. Thus, if we were to assume that such coherence was sufficient for justification, we should have to admit that any contingent statement a person is justified in accepting is such that he is also justified in accepting the denial of that statement.

Logical coherence is not, moreover, necessary for justification. Take any two observation statements describing observations of different and unrelated objects. Neither of these necessarily implies the other, but we can be justified in accepting both of them. Similarly, consider any two laws, one about stars and the other about mice. These may also be such that neither necessarily implies the other, but we may be justified in accepting both of them. The two laws or the two observation statements may be related in some way: they may be consequences of some more general law of nature, but they are not necessary consequences of each other.

These objections are decisive. The question is, How can a coherence theory avoid such difficulties? First, we must keep distinct the two questions

raised above, namely, what is coherence? and, second, what kind of a system is required? We shall not obtain a satisfactory coherence theory of justification by answering only the first question. In defending a belief or knowledge claim by arguing that it coheres with certain other beliefs, we must be prepared to explain why coherence with those beliefs provides justification. Hence, to articulate a satisfactory coherence theory, we must answer the second question as well. We must say what kind of system provides justification for those beliefs that cohere with the system.

Coherence As Explanation

A coherence theory of justification may affirm that the kind of coherence required for justification is explanatory coherence. Wilfrid Sellars first propounded such a view with Gilbert Harman, Bruce Aune and William Lycan following him. Harman has argued at some length that whether a belief is justified depends on the way in which it fits into the best overall explanatory account.[3] The question of whether a belief is justified cannot be decided, according to such a theory, in isolation from a system of beliefs. It is in relation to other beliefs belonging to a system of beliefs that the justification of a belief must be decided. Moreover, the system of beliefs determining justification must be one in which we explain as much as we can and leave as little unexplained as we must. A system having a maximum of explanatory coherence confers justification on beliefs within it.

If the kind of coherence required for justification is explanatory, then it is the function of a belief in explanation that justifies it. There are two ways in which a belief can so function. It can either explain or be part of what explains something, or it can be explained or be part of what is explained. To have explanatory coherence, one must both have something to explain and something to explain it.

Bertrand Russell once remarked that, though we do not know of the existence of physical objects, we may reasonably infer the existence of such objects because the hypothesis of their existence is the simplest and best explanation of why we experience the sense data we do.[4] A defender of the explanatory coherence theory of justification could reply that Russell does not go as far as explanation would warrant. The hypothesis that physical objects exist is such a good explanation of our experience of the sense data in question that we are justified in accepting and claiming to know of the existence of such objects. The traditional problem of the justification of perceptual claims on the basis of sense-data statements appears solved by the explanatory coherence theory.

Moreover, the problem of the justification of our claims about the mental states of others seems amenable to comparable treatment. If I see a man behaving just as I would were I in a certain mental state, then, one could ar-

gue, the best explanation I have for why he behaves that way is that he is in that mental state. Suppose, for example, that I see an injured man before me writhing, moaning, and otherwise behaving as I know I would if I were experiencing intense pain. The best explanation for why he behaves as he does is that he is feeling pain. To see that this is so, consider the problems one encounters with any hypothesis denying that the man is in pain. First, I must explain why the man is behaving in this way if he does not feel pain. Even if, however, my hypothesis does explain this, to obtain a satisfactory overall explanatory account, I must explain more. In explaining his behavior in some alternative way, I shall either assume that, though this man feels no pain, others in such circumstances would, in which case I must explain why this man does not feel pain when others would. Otherwise, I shall assume that others generally fail to feel pain in such circumstances, in which case I must explain why I do feel pain when others do not. In either case, I am left with an unsolved explanatory problem that would be avoided by hypothesizing that others generally, the man in question included, feel pain as I do under these conditions. From the standpoint of overall explanatory coherence, the latter hypothesis is obviously advantageous.[5]

As we proceed from perceptual claims and claims about the mental states of others to statements about distant times and places and, finally, to statements about theoretical states and objects, the appeal to explanation becomes more obvious and familiar. We might think it odd to justify the claim that we see our bodies or that our friends are suffering by arguing that it is best from the standpoint of explanation to suppose that these things are so, but it is commonplace to argue that hypotheses about the past, the physically remote, and the theoretically unobservable are justified by the way they explain what we seek to understand.

On the Justification of What Is Explained

The thesis that hypotheses are justified because of what they explain is most plausible, but how are we justified in accepting those things that are explained? If we claim that what is explained consists of basic facts and beliefs, we shall merely appeal to explanation to justify the inference from basic beliefs to nonbasic ones and adopt an explanatory version of the foundation theory. We are now considering a more radical departure from the foundation theory. According to the coherence theory under consideration, there are no basic beliefs. All beliefs are justified by their explanatory role. To explain, however, one must have something to explain as well as a hypothesis to explain it. What justifies those beliefs that provide the matter to explain?

The answer is that if some beliefs are justified because of what they explain, others are justified because they are explained. Moreover, it is plausi-

ble to suppose that some beliefs are justified because they are so well explained. Suppose I look at a streak in a cloud chamber and conjecture that the streak is the path of an alpha particle. If I do not understand how an alpha particle could make such a streak, I may not be justified in my belief. Once it is explained to me how the alpha particle causes condensation, I may become justified in accepting that the streak is the path of the particle. It is no rare event in science or everyday life to have some doubt concerning a fact removed by some explanation of it. Such explanations may change dubious beliefs into justified ones.

Moreover, a belief may be justified *both* because it explains *and* because it is explained. That a chair supports me may explain why, in my present posture, I do not fall to the floor; and that the chair supports me may be explained, given my position on it, by the rigidity, and so forth, of the chair. Similarly, the path of the alpha particle may explain why we see what we do and may be explained within atomic theory. The same belief may be both explaining and explained, and it may derive justification from both roles. It is those beliefs that both explain and are explained whose justification seems most adequate. Indeed, explained unexplainers, such as sense-data statements, have been epistemically controversial, as have unexplained explainers, such as statements concerning the supernatural. Recently, philosophers of empiricist leanings have tended to construe the fundamental empirical statements as perceptual claims concerning physical objects rather than reports concerning sense data. The underlying reason for this tendency may be an unrecognized desire to settle on some empirical statements that are *both* explained and explanatory. Perceptual statements both explain sense experience and are explained by theories of perception.

Explanatory Coherence and Justification: An Analysis

The foregoing considerations substantiate the suggestion that justified beliefs are ones that explain or are explained, or both. Explanatory coherence thus appears to determine justification. We shall now attempt to offer a precise analysis of such justification. Let us reconsider the formula for coherence theories introduced earlier.

> *S* is justified in accepting that *p* if and only if the belief that *p* coheres with other beliefs belonging to a system C of beliefs of kind *k*.

To offer a coherence theory of justification, we must offer an account of coherence and of the kind of system with which a belief must cohere. Let us first consider the question of what kind of system of beliefs is required. Sellars suggests that our choice of a system should be one that yields a maxi-

mum of explanatory coherence, but a problem of interpretation arises immediately. A number of systems of beliefs compete for the status of having a maximum of explanatory coherence, and some of these systems might be ones that any given person could hardly conceive. Are we to require that a person's belief cohere with a system of beliefs of which he could not conceive in order for the person to be justified in what he believes?

One answer is that the required system be the one with a maximum of explanatory coherence of all those of which *S* could conceive. The other alternative is to require that his belief must cohere with that system having a maximum of explanatory coherence, whether he could conceive of it or not. Both answers present difficulties. One problem with the first answer is that a person might turn out to be justified in accepting something because of his inability to conceive of the system having a maximum of explanatory coherence with which his belief fails to cohere. One drawback of the second answer is that according to it a person might be justified in accepting something, even though it fails to cohere with systems he understands: his belief may cohere with a system of beliefs having a maximum of explanatory coherence that he is unable to comprehend and that is opaque to his understanding. Of the two difficulties, the latter appears the more severe. Hence, we shall suppose that the system of kind *k* is the one having a maximum of explanatory coherence among those systems of beliefs understood by *S*. We shall be able to elucidate further the concept of *maximal* explanatory coherence when we have clarified the notion of coherence.

Explanatory Coherence

Now let us consider the concept of coherence. One ingredient that has been assumed to be essential to coherence with a system is consistency with the system. We shall find reason to reconsider this assumption below, but it is a standard assumption that a belief fails to cohere with other beliefs when it is logically inconsistent with them. Consistency, even if necessary for coherence, as it is here assumed to be, is not sufficient for coherence, however, when the kind of coherence required is explanatory. To explicate this kind of coherence, we shall take the concepts of explanation and of *better* explanation as primitive, that is, undefined. It is agreed that these concepts are themselves in need of clarification. We shall consider the problems surrounding such clarification subsequently.

To cohere with the beliefs belonging to a system, a belief must fill an explanatory role, but what sort? It would be too restrictive to require that the belief explain or be explained by *all* beliefs belonging to the system. It may only explain or be explained by some beliefs in the system. Should we, therefore, say that a belief coheres with a system of beliefs if and only if it is consistent with the system and either explains or is explained in relation to

the system? No. A general belief may explain some belief within a system of the required sort, but fail to be justified because some other general belief that contradicts the first explains that belief better. Two contradictory general statements may each explain what is believed to be a fact, when one explains better than the other. Assuming a person cannot be justified in accepting both hypotheses, some additional restriction is needed.

We must require that a belief cohering with a system either explain or be explained in relation to the system better than anything that contradicts it. Contradiction must be made relative to the system. Two mutually consistent statements may be such that a system of beliefs entails that they cannot both be true. We shall speak of such beliefs contradicting each other and thus employ a relativized concept of contradiction. With this stipulation, the preceding problem is easily solved. A belief coheres with a system of beliefs if and only if the belief is consistent with the system and either explains something in relation to the system not explained better by any belief that contradicts it, or the belief is better explained by something in relation to the system, and nothing that contradicts it is explained better.

We thus obtain the sought-after notion of coherence needed to provide a coherence theory of justification as follows:

> *S* is justified in accepting that *p* if and only if the belief of *S* that *p* is consistent with that system *C* of beliefs having a maximum of explanatory coherence among those systems of beliefs understood by *S*, and the belief that *p* either explains something relative to *C* that is not explained better by anything which contradicts *p* or the belief that *p* is explained by something relative to *C* and nothing which contradicts it is explained better relative to *C*.

Let us now reconsider the concept of a system having a maximum of explanatory coherence. The preceding discussion suggests the way to elucidate this concept. If the beliefs belonging to one system explain better and are better explained than the beliefs belonging to a second system, then the first system has greater explanatory coherence than the second. A system *C1* has greater explanatory coherence than *C2* if and only if *C1* is logically consistent and *C2* is not, or both are consistent but more is explained in *C1* than *C2*, or both explain the same things but some things are explained better in *C1* than *C2*. We then adopt the following analysis of maximal explanatory coherence, which provides the sort of system of beliefs we sought at the outset:

> A system *C* has a maximum of explanatory coherence among those systems of beliefs understood by *S* if and only if there is no system having greater explanatory coherence among those systems.

This condition together with the preceding one constitutes a theory of justification in terms of explanatory coherence in which we have taken for granted the concepts of explanation, better explanation, and the usual logical notions.

On Explanation

The conception of explanation is, unfortunately, so interwoven with epistemic notions that we could not expect to explicate the idea of one explanation being better than another without at least covertly appealing to some epistemic notion. For example, one explanation is often said to be better than another solely because the first is more likely to be true from what we *know* than the second. Such considerations lead us in a small circle.

This difficulty can best be elaborated if we consider the concept of explanation *simpliciter*. There is an immense literature on this topic of considerable linguistic and formal sophistication. This literature illustrates most clearly the futility of hoping to find an explication of explanation to which we can fruitfully appeal in our articulation of the explanatory coherence theory. Consider first the deductive model of explanation admirably articulated by Carl Hempel.[6] With various refinements, this model of explanation tells us that F is explained by a statement of boundary conditions B and law L if and only if F is deducible from B and L in such a way that B and L are both essential to the deduction. Such analyses are wont to lead to implausible conclusions, the most notable of which is that almost any law can be used to explain almost any statement.[7] Moreover, the qualifications needed to eliminate such untoward consequences often appear to be entirely ad hoc. The more important objections to such analyses from the standpoint of the explanatory coherence theory rest on counterexamples.

Consider the following example, which is a modification of one proposed by Sylvain Bromberger.[8] Imagine that I am standing with my toe next to a mouse that is three feet from a four-foot-high flagpole with an owl sitting on top. From this information concerning boundary conditions, and the Pythagorean theorem, which we here construe as an empirical law, we can deduce that the mouse is five feet from the owl. Moreover, all the premises are essential to the derivation. Thus, in the proposed analysis, the boundary conditions, together with the law, explain why the mouse is five feet from the owl. Nonetheless, this deduction does not explain why the mouse is at that distance from the owl at all. If you have any doubts about whether this is an explanation, imagine that you *know* that the distance from the top of the flagpole to where you stand is five feet and that you have asked why the mouse is five feet from the owl. An answer to this question based on the boundary conditions cited and the Pythagorean theorem would not be explanatory. Receiving such an answer, you would, perhaps, apprehend how to deduce that the mouse is five feet from the owl from some premises, but

those premises do not explain why the mouse is five feet from the owl. And, moreover, the matter requires explanation—owls eat mice!

It is easy to see, moreover, that arbitrary applications of the Pythagorean theorem could be used to explain why any distance between any two objects is what it is. Just draw a right triangle with the distance between the two objects constituting the hypotenuse, measure the sides, perform the calculation, and you have explained the distance between the two objects according to the analysis. Surely, that does not explain the distance between the two objects. You knew what the distance was prior to drawing the triangle, and whatever you accomplished by drawing the triangle, measuring the sides and performing the calculation, does not explain why the distance between the two objects is what it is.

An Epistemic Analysis of Explanation

The sort of amendment required, according to Bromberger, is epistemic. An explanation supplies the right answer to a why-question, when the person to whom the matter is explained does not know the answer to the why-question and, indeed, would rule out any answer she could think of on the basis of what she does know. Such a person lacks understanding, and the understanding lacking is supplied by the explanation. These considerations lead Bromberger to offer an analysis of explanation consisting of an explication of sentences of the form SEBW, where *S* and *B* take expressions referring to persons as values, *E* takes some form of the verb 'to explain,' and *W* takes some question.[9] Thus, one instance of the formula would be as follows: Hempel explained to Lehrer why the mouse is five feet from the owl. One truth condition of this sentence is that Lehrer at first does not understand or know why the mouse is five feet from the owl. A second truth condition is that what Hempel communicates to Lehrer gives Lehrer knowledge of why the mouse is five feet from the owl.

The sort of analysis of explanation that Bromberger offers, however plausible or controversial it might be, cannot be exploited here without rendering the explanatory coherence analysis of knowledge immediately circular. Knowledge would be analyzed in terms of explanatory coherence, which would be analyzed in terms of explanation, which would be analyzed in terms of knowledge. Moreover, if we assume that Bromberger is correct (or very nearly correct) in his analysis, then it seems reasonable to conclude that if an analysis of knowledge is based on the concept of explanation, the latter concept should be taken as primitive in our analysis. This is the proper moral of the story.

Hempel has a reply to the preceding objection that we shall consider briefly.[10] He argues that the conception of explanation appealed to above, being relative to a subject and what he knows, is not the one he was attempting to explicate. The conception of explanation that Hempel claims

to be explicating is an objective logical relation between the law and what is subsumed under it. The question is whether the objective logical relation between a law and what is subsumed under it is a relation of explanation. For the purpose of explicating the logical structure of scientific theories, laws, and singular statements subsumed under them, it may not matter whether or not the relation of subsumption is that of explanation. For our purposes, however, it is crucial. The sort of counterexample considered above is conclusive here: the subsumption relation may fail to be explanatory and the explanatory coherence theory requires that justification be explanatory.

The preceding remarks are not offered as a refutation of Hempel's claim. They are, instead, a defense of our strategy of taking as primitive the conception of explanation, and of one explanation being better than another, in our discussion of the explanatory coherence theory. If we attempt to explicate the relation of explanation as an objective deductive relation of subsumption, we shall find that the relation fails to explicate why the premises of the deduction explain the conclusion. As we noted in the example of the owl and the mouse, sometimes the deduction is nonexplanatory. To distinguish adequately deductions that are explanatory from ones that are not, and to explicate what makes some deductions explanatory, we would have to appeal, as Bromberger contends, to epistemic considerations—to what we do and do not know. Such an appeal would render the analysis of knowledge circular. Instead, we shall take our explanatory conceptions as primitive. No thoroughly satisfactory nonepistemic analysis of explanation has been proposed, and, consequently, our remarks concerning theories of justification based on the concept of explanatory coherence must rely on an undefined notion of explanation. Yet we can find arguments against the explanatory form of the coherence theory strong enough to warrant abandoning it.

Objections and Replies to Coherence As Explanation

The first problem raised by our explanatory theory of justification concerns comparing systems with respect to explanatory coherence. Our theory tells us that one system has greater explanatory coherence than a second if the first leaves less unexplained or explains better what it does explain than does the second. Even so, one system may leave less unexplained and explain better what it does explain by containing less to be explained. One system may admit statements of unexplained facts that the other excludes. To reduce what is unexplained, one may refuse to concede the truth of those statements that need explanation. Explanation involves those statements that do the explaining, on the one hand, and those that describe

what is to be explained, on the other. One can increase the explanatory coherence of a system either by adding statements that explain or by subtracting statements to be explained. The method of increasing the explanatory coherence of a system by decreasing what is to be explained must be limited. Otherwise, we may obtain a maximum of coherence only by securing a minimum of content.

The foregoing remarks may be illustrated with a very simple formal example. Compare any system of beliefs within science to the following. Take a language with one observation predicate '*O*' and one theoretical predicate '*T*.' Then adopt a system affirming that everything is *T* and that everything *T* is *O*, hence, that everything is *O*. Let the system contain only these sentences. We can now get a maximum of coherence by adding just those observation sentences to our system that fit with our one empirical law. More concretely, if we wish to have the law affirming that all dragons breathe fire, we may then add the 'observation' sentences that object 1 is a fire-breathing dragon, object 2 is a fire-breathing dragon, and so forth. The coherence between the law and observation statement will be perfect, and the absurdity of the system will be manifest. To avoid this sort of implausibility, philosophers have imposed further limitations on what kinds of statements may belong to a justificatory system.

Explanatory Coherence and Observation Statements

Both W. V. O. Quine and W. F. Sellars suggest conditioned responses as one determinant of whether a statement is epistemically qualified.[11] Of course, this amounts to abandoning the theory of justification under consideration, for whether we are justified in accepting some observation statement to be true will then depend not only on its explanatory coherence with other statements but also on the existence of certain patterns of conditioned responses to sensory stimuli.

Quine and Sellars also advance theories of meaning according to which the meaning of terms and statements depends on the relations of those terms and statements to other terms and statements. Sellars would not identify the meaning of an observation statement with the pattern of conditioned responses in terms of which one responds with such a statement to sensory stimuli. Nevertheless, both authors consider such patterns to constitute the link between language and sensory experience. Hence, whether we are justified in accepting some observation statement to be true depends on how that statement is linked to sensory experience by such patterns. These patterns accordingly constitute some restraint on the way in which we may eliminate observation statements from the system to save ourselves explanatory labor.

Observation Statements and Conditioned Responses

Clearly, some amendment of the explanatory coherence theory is needed to preserve an explanatory base. Let us consider whether the present modification yields a satisfactory theory of justification. Consider the view that what makes a person justified in accepting some observation statement to be true depends, at least in part, on certain patterns of conditioned responses associated with the sentence. Of course, *action,* and not *belief,* is usually required as a response to episodes of belief acquisition. There is a defect in this proposal that is easy to appreciate and infects sophisticated modifications—a person may be conditioned to respond with erroneous beliefs.

Experiments in regard to perceptual beliefs concerning the size of coins show that a poor person will respond with erroneous beliefs much more frequently than one who is not. Let the experiment be one in which a person is shown a coin, then is shown a disc, and is asked to report whether they are the same size or whether one is larger than the other. The poor person will frequently judge the coin to be larger than it is. Is she justified in her belief? Of course not. What this shows is that conditioned responses can regularly produce erroneous as well as correct belief. Conditioning in and by itself is neutral with respect to truth and error.

The preceding remarks are not intended to refute the proposal that a person might be fortunate enough to be justified in accepting some observation statement whenever her belief is a conditioned response to a certain kind of stimulus. It may be true, just as it may be true that she is justified in accepting an observation statement whenever she is in a brain state of some special kind. Even if such beliefs happen to be justified, however, it is not the conditioning or the brain state that *makes* them justified. If people happen to be so conditioned that what they believe is justified, that is fortunate. Still, they could equally well have been so conditioned that what they believe would not be justified. It may be that I am conditioned to believe that an object is red when I am confronted with a red object in certain circumstances, but I could equally well have been conditioned to believe that such an object is yellow. The latter belief would not have been justified. What makes the belief justified is not the conditioning, even if the justified belief is a response to a conditioned stimulus.

The foregoing argument applies against any theory affirming that what makes a belief justified ever depends on the belief being a conditioned response to a stimulus of a certain kind. What a person is conditioned to believe is one thing, and what she is justified in accepting is another, even if the two happen to coincide. Of course, we condition a child to have be-

liefs that we think are justified and discourage beliefs we think are unjustified. Nevertheless, it is not her being so conditioned that makes her belief justified.

Observation and Spontaneity

Another attempt to solve the problem of accounting for observation, suggested by Laurence BonJour, would be simply to impose an observation requirement to the effect that various beliefs that occur spontaneously are reliable or likely to be true.[12] There are two objections to this technique. First of all, it does not seem to be the spontaneity of the beliefs that accounts for their justification but rather their content. It is because observation beliefs are about what we observe that they are justified, not because they are spontaneous. What a person spontaneously believes is one thing, but what she is justified in believing is quite another. A person given to having spontaneous beliefs about demons and monsters would not be justified in such beliefs. The most critical objection to imposing such a requirement, however, is that it must be either arbitrary or unnecessary. If our system of beliefs gives us no justification for accepting that beliefs about what we observe are reliable, then the requirement is arbitrary. If, on the other hand, our system of beliefs does give us a justification for accepting that such beliefs are reliable, then the requirement is unnecessary.

Observation and Natural Selection

Yet another way of saving observation statements, by appeal to the theory of natural selection, is equally faulty for similar reasons. To argue that beliefs about what we observe must be justified because they have survival value in the process of natural selection will leave one epistemically bankrupt. First, the form of survival theory that currently appears most tenable is one recognizing that many factors bear little weight in the struggle for survival and, consequently, may be retained even though they have almost no survival value. Hence, one cannot argue directly from the existence of beliefs to their survival value. Second, and more important, even if this inference is allowed, the epistemic leap to the conclusion that such beliefs are justified is totally unwarranted. Beliefs that are neither true nor justified may have considerable survival value. Perhaps the truth would destroy us.

An Ethical Analogy

One final argument. Consider briefly the parallel between ethics and epistemology.[13] R. M. Chisholm, following C. I. Lewis, has argued that a theory of justification provides a criterion of evidence and justification just as a theory of ethics provides criteria of right and wrong. Imagine a person arguing that an action he performed was right because he was conditioned to

perform that action or because the performance of such actions has not been extinguished through the process of natural selection. The latter contention is absurd on the face of it.

If we believed that a person was conditioned to perform a certain action, we might conclude that he could not help but perform it, that he was responding to a kind of compulsion, and thus refuse to condemn him. However, if the action was one intentionally aimed at producing wanton pain and suffering in others, we would not condone the action as right. The action was not right even if the person could not help but perform it. Similarly, if a person is conditioned to accept something that he is not justified in accepting, it hardly follows that he is justified. The claim that a person is justified in accepting something because it is a conditioned response to sensory stimulation is no better warranted than the claim that a person is right in performing an action because it is such a response. Conditioning fails to justify our beliefs. Justification emanates from another source.

We conclude that the appeal to conditioned responses, however interesting psychologically, will not suffice as the basis of a supply of justified observation statements to be explained within an explanatory system. Moreover, as we noticed earlier, the appeal to conditioned responses amounts to introducing an additional factor into the explanatory coherence theory. Is there any way of preventing the wholesale depletion of observation statements from our system of beliefs without abandoning the theory of justification as explanatory coherence? In fact, there is a way.

Self-Explanatory Beliefs

One could maintain that observation statements are self-explanatory and hence that a gain of explanatory coherence results from the inclusion of such statements within the system.[14] How can a statement be self-explanatory? When the truth of *p* explains why the person believes that *p*. For example, suppose I believe I see blood on my shoe. How is my belief to be explained? One explanation of why I *believe* that I *see* blood on my shoe is that I *do* see blood on my shoe. According to Bromberger, this explanation appeals to an *exceptive* principle.[15] In answering the question of why I believe that I see blood on my shoe, we are presupposing a principle affirming that no one believes that he sees blood on his shoe *except* when he does see blood on his shoe, or when he incorrectly takes what is on his shoe for blood, or when he is hallucinating, and so forth. If I believe that none of the other alternatives is correct and such beliefs cohere with my system of beliefs, then the statement that I see blood on my shoe explains, at least in part, why I believe that I see blood on my shoe. As a consequence, my belief is justified.

The preceding argument shows how a belief could be justified by being self-explanatory. If what is believed is true, then the truth of the belief explains, at least in part, the existence of the belief. Of course, a fuller explanation should be forthcoming, for instance, one explaining how I happen to see blood on my shoe, perhaps because my bruised foot is bleeding, and so forth. Though the explanation is incomplete, it is acceptable as far as it goes. Moreover, beliefs most plausibly taken to be self-explanatory in this way seem to coincide with perceptual beliefs. Hence, this sort of justification promises to provide a base of justified observation statements. Memory statements seem amenable to comparable treatment.

Before turning to a critical examination of this theory, we shall note its virtues. It offers the possibility of providing justified perceptual beliefs within the context of an explanatory coherence theory without dragging in some nonexplanatory feature to account for their justification. Such beliefs are justified because the truth of the belief explains the existence of the belief, that is, the statement that a person believes what he does is explained by the statement that what he believes is true. Moreover, this explanation depends on the *system* of beliefs a person has, and, consequently, on other beliefs in that system, for example, those that exclude alternative explanations for the existence of the belief. The self-explanatory justification is, therefore, not a form of self-justification. The self-explanatory character of perceptual beliefs depends on explanatory coherence within a system of beliefs.

Self-Explanation: An Evaluation

The proposal that justification can be obtained through self-explanation in a system of beliefs, though promising, evokes criticism appropriate to the explanatory coherence theory as a whole. It may be doubted whether the purported self-explanation is explanation at all, and it may be affirmed that the justification obtained does not depend on explanation. We shall examine these objections as they apply to the theory of justification through self-explanation and then to the more general form of the theory of justification by explanation.

First, it might be objected that the general principles involved in self-explanation, *exceptive* principles, are trivially true and, consequently, no explanation can be based on them. The principle that no one believes he sees something *except* when he does see it or when he erroneously takes something else for it or when he is hallucinating, and so forth, has the appearance of a tautology. It tells us no more than that no one believes that he sees something except when he sees it or when he erroneously believes that he sees it. This, it might be objected, is not a principle of explanation but is the barest of tautologies. Moreover, people sometimes see things they do not believe they see just as they sometimes believe they see things when they do

not. If the spot on my shoe is nail polish, not blood, then I may not believe that I see a spot of nail polish on my shoe though I do, and I may believe that I see blood on my shoe though I do not. Thus, we lack any explanatory law, exceptive or otherwise, to provide an explanatory link between what a person believes she sees and her seeing that object.

Second, it might be contended that if a person believes that she sees something immediately before herself, the existence of this belief justifies us in accepting that she sees it, at least when one has no reason to doubt that she sees what she believes she does. If her belief justifies *us* in concluding that she sees the object in question, then her belief must also justify her. Thus, the example illustrates a self-justified belief, but it does not depend on explanatory considerations. That a person believes that she sees something provides justification for concluding that she does see it without explaining or being explained by any other belief or statement. That is the objection.

One might defend the explanatory coherence theory against such objections either by rejecting the doctrine of self-explanatory beliefs or by maintaining that such explanation is genuine. We cannot offer any decisive argument against the possibility of sustaining these alternatives, but neither seems tenable. Of course, these remarks are no defense of a foundation theory against a coherence theory. The way in which perceptual beliefs cohere with a system of beliefs may render them justified even though the coherence is not explanatory. Coherence may be explicated in some other manner, and so we now turn to other criticisms of the explanatory coherence theory.

Justification Without Explanation: Some Examples

The first example of a justified belief whose justification does not depend on explanatory considerations was presented above. It is the example in which a person deduces from the Pythagorean theorem and boundary conditions that the mouse is five feet from the owl, even though he has no explanation of why this is so. The belief is justified, but the justification of the belief does not depend on explanatory relations. It is enough that the person knows the Pythagorean theorem, the distance to the pole, and the height of the pole and then deduces the conclusion. He is thus justified in his belief that the mouse is five feet from the owl, even if he has no idea how to explain that nor any idea about how to explain anything else in terms of that belief.

For a second example, suppose that David Hume in eighteenth century Edinburgh sees a dead man before him. If asked whether the dead man was sexually conceived, Hume would reply that he was and would be justified

in his belief because he would be justified in accepting that all who die are conceived. Death, however, does not explain conception any more than conception explains death.[16] The constant conjunction Hume observed between conception and death did not indicate a causal connection. Conception does not cause death any more than death causes conception. Neither explains the other. Moreover, the observed conjunction will cease to hold in our century and is no law when some die who were laboratory artifacts. Hume was justified in accepting that the dead man was once sexually conceived, nonetheless.

These are two examples of justified beliefs whose justification does not depend on explanatory relations to other beliefs. They may be neither explained nor explanatory, but they are justified because they cohere, in some way independent of explanatory function, with the other beliefs within a system of beliefs. Because the beliefs that the mouse is five feet from the owl and that the dead man was once conceived cohere with other beliefs, they are justified, but the coherence is not explanatory.

A defender of the explanatory coherence theory could reply to these objections that we have ignored the way in which the conclusions and premises in question function in the overall explanatory system. She might argue that those beliefs are only justified because of the explanatory relations of those beliefs within an overall system having a maximum of explanatory coherence. She could also claim that the Pythagorean theorem and the general principle concerning how people come to exist are themselves justified because of their systematic explanatory role. Finally, she could say that such general beliefs within an overall system having a maximum of explanatory coherence are what make our conclusions justified.

It is difficult to comment on this reply without indulging in simple counterassertion. However, with some imagination we may, I believe, construct something of an argument. Imagine a group of people who, perhaps because of their religious beliefs, meticulously avoid asking for explanations of what they observe. They are anti-explanationists. Anti-explanationists ask not why or how things happen but are content to observe the way things happen and rely on such observations without seeking explanations. They pride themselves in their intellectual humility. Such people might arrive at the Pythagorean theorem from observation. They may not inquire as to why it is true and they may not have deduced it from more general axioms. Nonetheless, they might be justified in accepting what they derive from it, for example, that the mouse is five feet from the owl, whether or not the theorem or the conclusion derived from it contributes to the explanatory coherence of some overall system of beliefs. It would be most peculiar to affirm that what made them justified in accepting what they did on these matters was the explanatory role of such beliefs within the system of their beliefs. They might be wholly oblivious to such explanatory virtues,

and, indeed, would be indifferent or perhaps even hostile to receiving suggestions concerning the explanatory merits of what they believed. What makes them justified in accepting what they do is connected with the way in which these beliefs cohere with a system of beliefs they have, but the coherence involved is not explanatory. Thus, explanatory coherence is not necessary for justification.

Explanationism and Opacity

It might be objected that the anti-explanationists are simply ignorant of what *makes* their beliefs justified, namely, the explanatory coherence of their system of beliefs, even though they have no idea that this is so. This objection again confronts us with the opacity problem. If the anti-explanationists are oblivious to facts of explanation and fail to notice that one thing explains another, that some belief is explained, for example, then why should the fact that the belief is explained justify the belief for them when the fact goes unnoticed by them? If the feature of explanation is opaque to them, perhaps because of the aversion to it, then the presence of it will fail to justify them in their beliefs. Suppose, moreover, that their system of beliefs based on observation, empirical generalization, and deduction from those generalizations is not sufficient to ensure that anything is explained. It might, nevertheless, suffice to justify them in believing what they do because the system reveals the evidence and justification to them.

Some Final Objections: Weak Explanations and Competing Systems

There remain two related objections that illustrate some problems to be solved by any satisfactory form of the coherence theory, explanatory or not. First, suppose that some hypothesis provides a better explanation of other beliefs within a system having a maximum of explanatory coherence, even though the explanation is not fully adequate. Imagine, for example, that a man has been shot and that the maid is the prime suspect. Her fingerprints are on the gun and she admits the deed. Moreover, suppose that she has a motive. Nevertheless, imagine that she has never fired a gun previously, the spot from which she would have had to fire the gun was a good distance from the victim, and, moreover, there are footprints outside the window and in the room where the crime took place, made by boots that clearly were not owned by the maid. Even if the maid avows that she made the footprints with boots to turn suspicion away from herself and then destroyed the boots, we may have our doubts. The hypothesis that she shot the victim may be the best explanation because we can conceive of no better one. In this situation, we would not claim to know; there is too much in

doubt for that. And even if we do believe that the maid is the killer, we would not be justified in accepting this. We would not think we were that well justified, nor would others.

Following the explanatory theory offered above, we would be justified in accepting that the maid is the killer because it is the best explanation we have. This suggests that we must require that a hypothesis not only explain better than any alternative we can conceive but also that it be a comparatively good explanation, good enough so that we are justified in accepting it. More generally, beliefs must cohere in some comparatively *strong* way with other beliefs within a system for such coherence, however it is explicated, to yield justification.

Our final objection to the explanatory coherence theory is that it has a defect characteristic of coherence theories, to wit, inconsistent statements turn out to be justified. Two systems of beliefs may each have a maximum of explanatory coherence and yet be inconsistent with each other. There may be two or more systems of beliefs each having a maximum of explanatory coherence. Each may be such that no other consistent system of beliefs leaves less unexplained, and none explains what it does explain better. Consequently, a belief may cohere with one system of beliefs having a maximum of explanatory coherence while the contradictory of that belief coheres with another system of beliefs having a maximum of explanatory coherence. In the current account both beliefs would be justified.

One might attempt to meet this objection by (i) requiring that there be *one* system which is, from the standpoint of explanation, the best or by (ii) requiring that the concept of justification be made relative to a system. Both of these maneuvers fail. The first fails because we have no reason to believe that there is *one* best system from the standpoint of explanation. There are always conflicting theories concerning some aspect of experience that are equally satisfactory from the standpoint of explanation. Hence, if it is required that there be a best overall system before any belief is justified, we shall never be justified.

As for the second suggestion, that justification be made relative to a system of beliefs, there remain two objections. First, and perhaps most important, the question of whether a person is justified in accepting that p is not answered by the announcement that he is justified in accepting it relative to a system B. We must ask whether a person who is justified in his belief relative to system B is actually justified in his belief. In other words, is system B a system to which a man may legitimately appeal to justify his beliefs? If B is but one of a set of systems having maximal explanatory coherence that are inconsistent with each other, then we have no way of answering this question.

We are left with the problem of inconsistent *systems* of beliefs having a maximum of explanatory coherence and, consequently, inconsistent beliefs

being justified by such systems. It is interesting to notice that the very defect of idealistic coherence theories—the inconsistency of equally coherent theories—is also a defect of the theory of explanatory coherence. Moreover, the difficulty is not hard to discern. No relation between statements suffices for justification. In addition to relations between statements, some other feature must be an ingredient of justification.

Simplicity and Conservation

Some philosophers, such as Sellars, Quine, and Harman, for example, have appealed to the simplicity of the overall system to supply the needed additional ingredients.[17] Of two systems, both of which have a maximum of explanatory coherence, the simpler of the two is the one providing justification for beliefs within it. There are some objections to this strategy. First, simplicity is both obscure and complex. The complexity of simplicity results from the different ways in which one system can be simpler than another and from a certain stress between these modes of simplicity. One system may be simpler than another in terms of the postulates of the system, in terms of the basic concepts of the system, or in terms of the ontology of the system. We have at least postulational, conceptual, and ontological simplicity to consider, and these modes of simplicity may conflict. We sometimes purchase conceptual simplicity at the cost of multiplying entities in our ontology. Moreover, the notion of simplicity is hardly pellucid. It is difficult, even on intuitive grounds, to judge when one system is simpler than another. When a philosopher says his system is simpler than another, one may fairly suspect him of special pleading for the sort of system he prefers. Perhaps there is some common feature of such preferred systems. Or maybe such preferences are shaped by the cognitive fashions of the decade. No matter, we may reasonably doubt whether there is any sufficiently articulate conception of simplicity to which impartial appeal could be made in choosing between explanatory systems.

Even if we were to grant, however, that there is some serviceable conception of simplicity, this would fail to resolve the problem before us. There may be two systems that are not only maximal with respect to explanatory coherence but are also minimal with respect to complexity. If we have two systems that are equally coherent and equally simple, we shall have no way of deciding which system provides justification for the beliefs within it. Moreover, we actually complicate matters by introducing the concept of simplicity. Now we must balance simplicity against coherence when, for example, one system is slightly more coherent and leaves more unexplained, whereas the other is slightly simpler and presupposes a smaller ontology.

Finally, the appeal to simplicity exacerbates a problem we left unsolved above, namely, that we may justify beliefs by depleting a system of statements to be explained. By rejecting concepts and entities, we can obtain a simpler system as well as a more coherent one. If we seek both simplicity and coherence, we shall have the very strongest motive for rejecting observation statements for the purpose of reducing what needs to be explained, thereby obtaining greater explanatory coherence and simplicity. We again confront the sterile simplicity of a system confined to one theory, one law, and one set of confirming singular statements. Everything else may be disposed of hygienically to avoid explanatory untidiness and thus keep the system clean and neat.

The authors cited appeal to a principle of conservation in an effort to escape the unwanted diminishment of the system. Sellars stresses the need to conserve observation statements.[18] Quine and Harman refer to a principle of conservativeness or laziness in the general retention of beliefs.[19] If we apply their remarks to the problem before us—it is proposed that if two systems are equal in explanatory coherence and simplicity, and all others are less coherent and less simple—then that system provides justification for beliefs within it that conserves what we believe, at least among statements of a specified variety.

The primary problem with this proposal is simply that it is a principle of epistemic conservatism, a precept to conserve accepted opinion. Sometimes such a precept provides good counsel, but often it does not. The overthrow of accepted opinion and the dictates of common sense are often essential to epistemic advance. Moreover, an epistemic adventurer may arrive at beliefs that are not only new and revelatory but also better *justified* than those more comfortably held by others. The principle of the conservation of accepted opinion is a roadblock to inquiry and, consequently, it must be removed.

The preceding remarks are less than argument. Moreover, this principle of conservation, though wide of the mark, embodies at least one important insight, to wit, that whether a person is justified in accepting something depends on what she actually accepts, her system of actual acceptances. Indeed, the fact of acceptance itself—the subjective reality of positive evaluation—provides the basis for a satisfactory coherence theory of justification. Such a coherence theory of justification, based on the existence of acceptance and on the reasonableness of such acceptances, will render maxims of conservation and stability unnecessary and unwarranted. Such a theory based on what a person actually accepts contains within it an explication of the way in which shifts and changes of acceptance, however radical, bring with them changes in what a person is justified in accepting. In the next chapter we turn to the development of these ideas.

Summary

We have found three major reasons for rejecting the explanatory coherence theory of justification developed at the beginning of the chapter. First, the explanatory coherence of a system could be increased by decreasing what needs explanation. We thus reduce the problem of explanation by systematically denying the truth of those statements describing whatever is unexplained until we obtain a very simple system in which everything is perfectly explained because there is almost nothing to explain. No explanatory function or role of statements suffices to prevent this artificial manipulation of explanatory systems. Second, we found examples of statements and beliefs that were justified by general statements within a system, such as the conclusion derived from the Pythagorean theorem, quite independently of any explanatory role or function of such statements. Finally, systems may tie for the award of being the system with a maximum of explanatory coherence. A statement justified with respect to one such system is not justified with respect to another. Indeed, some statement inconsistent with the first may be justified in another equally maximal system.

All these difficulties spring from the same source. Having abandoned a foundation theory in which justification is built on self-justified basic beliefs, we are led by the explanatory coherence theory to build justification on the explanatory relations between statements. Such explanatory relations will not suffice, however. Explanatory relations between statements fail to pick out a *unique* set of justified beliefs, because we may, with sufficient imagination, concoct a myriad of different systems of statements in which such explanatory relations hold. Explanatory relationships *can* yield justifications—in this the theory is correct—but there must also be some other ingredient determining what needs to be explained in the first place. Here one might be tempted to waver and return to the foundation theory for a supply of basic beliefs in need of explanation. But that way is closed. We must proceed without a signpost guaranteeing the way to truth. There is nothing other than the coherence among our beliefs on which to rely.

The element needed to produce a sound coherence theory has been constantly before us. The goal of acceptance is to obtain truth and avoid error. This is why the objective of maximizing explanation is neither necessary nor sufficient for the sort of justification we have been seeking. We may aim at truth without aiming at explanation, and we may aim at explanation without aiming at truth.[20] We need not seek any guarantee for the truth of what we accept, nor need we appeal to explanatory relations among what we accept to provide a justification. A set of acceptances that arise from the quest to accept what is true and avoid accepting what is false can provide justification to target acceptances among its membership without appeal to explanation, simplicity, or conservation, whereas those that arise from an

interest in other matters may prove epistemically impotent. Let us consider what sort of justification may be obtained simply from fidelity to the goal of accepting something just in case it is true.

Introduction to the Literature

The traditional defender of the explanatory coherence theory is Wilfrid Sellars in *Science, Perception, and Reality.* Bruce Aune carried on the tradition in *Knowledge, Mind, and Nature.* Gilbert Harman has written two very readable books, *Thought* and, more recently, *Change in View.* The former is, perhaps, the most accessible formulation of the explanatory coherence theory. See also Alan Goldman's *Empirical Knowledge,* William Lycan's *Judgment and Justification,* and Jay Rosenberg's *Our World and Our Knowledge of It.* For a single article applying the explanatory coherence theory to the problem of other minds, see Paul Ziff's "The Simplicity of Other Minds."

Notes

1. See Robert Fogelin, *Evidence and Meaning* (New York: Humanities, 1967), 94–98.

2. They also conceive of coherence as truth. See Brand Blanshard's *The Nature of Thought* (London: Allen & Unwin, 1939), particularly vol. 2, chaps. 26–27, pp. 250–331.

3. See Wilfrid Sellars, "Some Reflections on Language Games," in *Science, Perception, and Reality* (London: Routledge & Kegan Paul, 1963), 321–58; and Gilbert Harman's chapter, "Induction," in *Induction, Acceptance, and Rational Belief,* ed. Marshall Swain (Dordrecht: Reidel, 1970), 3–99, as well as Harman's book *Thought* (Princeton: Princeton University Press, 1973). See also Bruce Aune, *Knowledge, Mind, and Nature* (New York: Random House, 1967); and William Lycan, *Judgment and Justification* (New York: Cambridge University Press, 1988). The present author defended a similar view in "Justification, Explanation, and Induction" in Swain, *Induction,* 100–33.

4. Bertrand Russell, *The Problems of Philosophy* (London: Holt, 1912), 22–26.

5. See Paul Ziff, "The Simplicity of Other Minds," in *New Readings in Philosophical Analysis,* ed. H. Feigl, W. F. Sellars, and K. Lehrer (New York: Appleton-Century-Crofts, 1972), 418–23.

6. Carl G. Hempel, "Studies in the Logic of Explanation," in *Aspects of Scientific Explanation* (New York: Free Press, 1965), 245–90.

7. See R. Eberle, D. Kaplan, and R. Montague, "Hempel and Oppenheim on Explanation," *Philosophy of Science* 28 (1961): 418–28.

8. Sylvain Bromberger, "Why-Questions," in *Mind and Cosmos,* ed. Robert Colodny (Pittsburgh: University of Pittsburgh Press, 1966), 105.

9. Bromberger, "An Approach to Explanation," in *Studies in Analytical Philosophy,* ed. Ronald Butler, 2d series (Oxford: Blackwell, 1965), 72–105.

10. Hempel, *Aspects of Scientific Explanation*, 425–28.

11. See Sellars's remarks on language entry transitions in "Some Reflections on Language Games," 321–58. It should be noted that Sellars explicitly denies that such conditioned responses are sufficient for establishing the meaning or justification of observation statements. Nevertheless, they do play a role in such justification. See Quine's remarks on stimulus meaning in *Word and Object* (Cambridge: MIT Press, 1960), chaps. 1–2.

12. Laurence BonJour, *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press, 1985), 141–46. A detailed statement and defense of coherentism regarding empirical knowledge is found in chaps. 1, 4, 6, 7, 9.

13. See the introduction to C. I. Lewis's *Mind and the World Order* (New York: Dover, 1929); and R. M. Chisholm, *Perceiving: A Philosophical Study* (Ithaca, N.Y.: Cornell University Press, 1957), chap. 3, "The Problem of 'the Criterion,'" 30–39.

14. See Alan Goldman, *Empirical Knowledge* (Berkeley: University of California, 1988).

15. Bromberger, "Why-Questions," 96–102.

16. This example was provided by Frederick Schick.

17. In articles and books cited above.

18. Sellars, "Some Reflections on Language Games," sec. 85, p. 356.

19. Quine in *Word and Object*, chap. 1, pp. 20–21; and Harman in *Thought*, 159.

20. The importance of the objective of truth-seeking in a coherence theory has been systematically emphasized by BonJour in *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press, 1985), whereas the opposite view that explanation suffices is defended by William Lycan, *Judgment and Justification*.

6

Internal Coherence and Personal Justification

JUSTIFICATION IS COHERENCE with a background system. In the preceding chapter, we considered a theory according to which justification consists of explanatory relations within a system of beliefs. The objectives of accepting what is true and avoiding accepting what is false are best served, according to that theory, by maximizing explanatory coherence. A principle can serve the purposes of truth, however, while lacking explanatory merit. The Pythagorean theorem suffices to obtain the truth about distances, whatever its explanatory limitations. We shall correct the defect in the explanatory coherence theory by giving truth its due. To this end, we emphasize acceptance as the central notion. It is what we accept in the interests of obtaining truth and avoiding error, our *acceptance system,* that constitutes the core of the relevant background system. Coherence with our background system is determined by what it is reasonable to accept based on this system. A concern for truth and nothing but the truth drives the engine of justification.

We shall begin with what we accept. From what we accept, we shall generate a notion of subjective or personal justification. Why begin with acceptance? Not because we have any guarantee of truth in that domain. We err about the character of our own mental states as we do about the external world. Indeed, our capacity for precise observation of the external world of objects and properties is more refined than our capacity for observing the

internal world of thoughts and sensations. We start with what we accept for lack of another alternative. One might protest that we should begin with experience, with the prick of sense. The stimulation of our senses raises the question of what we should accept, however, instead of answering it. Our senses may give rise to some representation or belief about what is transpiring in our sensory neighborhood. How are we to decide, though, whether what is suggested to us by our senses is true and accurate rather than false and illusory? We must consult information about the matter. What is this information? It is what we have accepted in the quest for truth. It is our background system of accepted information. The evaluation of all claims to truth, whether those of our senses, of reasoning, of memory, or of the testimony of others, must be based on our acceptance system, which includes our conception of the world and our access to it. There is no exit in evaluation from the circle of what we accept. Acceptance is the fuel for the engine of justification.

Acceptance and Belief Reconsidered

How does an acceptance system generate justification? It does so by telling us how reasonable it is to accept something in the quest for truth. Our acceptance system tells us it is more reasonable to accept one thing than another and more reasonable to accept something on one assumption than on another when we seek truth, that is, seek to accept something if and only if it is true. Consider the distinction between acceptance and belief that we propounded in Chapter 2. Many false ideas are presented to us in attractive ways and may, as a result, be believed when we know they are false. A politician may convince you of the truth of what he says when you know that he is untrustworthy. You know he will say whatever it takes to obtain your vote with practiced persuasiveness. His eyes radiate his ambition. But you want to believe him. You want to believe that the economy is strong and that you are economically secure, especially if the objective measures are alarming. He is warm, human, and comforting, whereas the data are cold, mathematical, and distressing. How can you resist? You do believe him, but you know that the economy is slipping.

How are we to account for this conflict between knowledge and belief? We are divided into separate systems. One is truth-seeking, and it contains what we accept in the interests of obtaining truth and avoiding error, of accepting something just in case it is true. The other system of belief is the yield of habit, instinct, and need. Often the two coincide. For the most part, what we believe is also something we accept in the interests of obtaining truth and avoiding error, and what we accept in this way is also something we believe. Sometimes the ways diverge, however. Sometimes the voice of truth speaks against the more ancient tongue of belief. We do not accept

what the politician tells us as the bona fide truth even if we cannot help but believe him.

Take another example. We look at the stars on a summer night and believe that they all exist somewhere far away from us in the heavens. As we look, we cannot help but believe that these bright objects now exist. Yet science tells us that some of them have long since disappeared and only the light traveling through space is reaching us after an astronomic delay. Science, not our eyes, is to be trusted, and so we do not accept what our eyes tell us. An ancient system of perceptual belief conflicts with the scientific system of acceptance. Traditional philosophers spoke of a conflict between reason and belief, whereas modern philosophers may be more inclined to speak of a conflict between a central system capable of ratiocination and some more automatic input system. Whether we prefer a traditional or modern construction, one system, the acceptance system, acts as judge of the other to obtain truth and avoid error.

Justification and Reasonable Acceptance

An acceptance system yields justification by informing us that it is more reasonable to accept some things than others, but how does it do this? The answer should be clear from what has gone before. An acceptance system tells us when we should trust our sources of information and when not, when we should trust our senses and when not, when we should trust our memory and when not, when we should trust the testimony of another and when not, when we should trust mathematical reasoning and when not, when we should trust some method of science and when not.

My acceptance system tells me that if I see what looks like a handbag on my dining table in the front room, it is more reasonable to trust my eyes and accept that it is a handbag than to accept that it is not. By contrast, if I see what looks like a handbag in a plastic case in an art museum with a label reading "Ceramic Object," my acceptance system tells me it is more reasonable for me to accept that the object is not a handbag but a ceramic sculpture of one. I may be wrong in both instances. My wife may have purchased a ceramic work of art and put it on our dining table, or the artist may have put a leather handbag in the plastic case to construct a work of conceptual art with a misleading label. My acceptance system is fallible, naturally, but it is the instrument I must use at this moment to decide what to accept on the basis of the information I now possess. In deciding whether to accept something or not at the present moment, reason requires the use of the relevant information I have accumulated in the quest for truth. That information is contained in my acceptance system.

My acceptance system changes in response to new data and further ratiocination, but at any moment it represents the outcome of my efforts, how-

ever brilliant or ineffectual, to distinguish truth from error. I may be justly criticized for not having done better in sorting truth from error in the past, but I cannot be faulted for judging now on the basis of my present acceptance system. I confront the question of whether or not to accept some information that I receive—that the economy is improving or that a star exists. My acceptance system answers the question by telling me how reasonable it is to accept the information in comparison to other competing considerations. If, on the basis of my acceptance system, the information is more trustworthy considered in terms of source and circumstance than conflicting or undermining objections, then it is more reasonable for me to accept the information on the basis of that system because of the way it coheres with that system. That is the way coherence yields justification.

Justification, Reasonableness, and Coherence

We are now in a position to give an account of coherence and justification based on the notion of a background system. We began with the schema

> *S* is justified in accepting that *p* at *t* if and only if *p* coheres with system *X* of *S* at *t*

and noted that it was necessary to specify a system and a relation of coherence in order to complete the account. We have indicated that we are going to begin by giving an account of personal or subjective justification. The core of the appropriate system for explicating such justification is the acceptance system of *S* at *t*. The role of the background system is to provide us with an evaluation of the reasonableness of what we accept, but there are materials of evaluation surrounding the core of acceptance that we must add to obtain the needed account of the background system of evaluation.

A background system providing an evaluation of the reasonableness of what we accept, which we shall call the *evaluation system* of the person, consists of the core acceptance system and related systems concerned with acceptance. The first such system to consider is our *preferences* concerning what we accept. If a person prefers accepting *p* to accepting *q* in the interests of accepting what is true and avoiding accepting what is false and the person is reasonable in her preference, then it is more reasonable for the person to accept *p* than to accept *q* in terms of those interests. A person having such a preference may, however, accept neither *p* nor *q*, perhaps because of some doubts about both or simply because of cognitive limitations, those of intellectual focus or memory, for example.

A scientifically minded person who doubts the existence of demons may not accept anything concerning the evil demon of Descartes, having never considered the matter or, perhaps, lacking any interest in Descartes and his reflections. Such a person might, nevertheless, prefer accepting that there is no such demon to accepting that there is in the interests of accepting what is true and avoiding accepting what is false in the way in which a diabetic person prefers eating sopapillas without sugar to eating sopapillas with sugar in the interest of good health even though she had never considered eating sopapillas. Thus, we shall include the preferences that a person has concerning what to accept in the interests of accepting what is true and avoiding accepting what is false in the evaluation system of a person because of the relevance they have to evaluations of what it is reasonable for a person to accept.

A second system to be added to the evaluation system of a person is the system consisting of reasonings from acceptances to further acceptances as conclusions. How a person reasons from premises she accepts to conclusions she accepts is germane to evaluations of what it is reasonable for a person to accept. If a person reasons cogently from premises she reasonably accepts to further conclusions, then she reasonably accepts those conclusions. Thus, the reasonableness of acceptance is elucidated by how a person reasons. Reasoning uses and extends acceptance, but it is not reducible to it. It is one thing to accept both that q is true if p is true and to accept that p is true on the one hand, and another to reason cogently from accepting these things to accepting q. Reasoning to a conclusion goes beyond accepting the premises of the reasoning, and is not reducible to it. So, we must add the reasonings of a person to the evaluation system of person employed to evaluate the reasonableness of acceptance.

The evaluation system of a person consists of what the person accepts, what the person prefers concerning acceptance, and how the person reasons concerning acceptance. This is the system that determines the reasonableness of an acceptance. We may, therefore, fill in the reference to a system in our analysis of coherence, noting at the same time the restriction to an account of personal justification, as follows:

> S is personally justified in accepting that p at t if and only if p coheres with the evaluation system of S at t.

Personal justification will provide us with a first component in an adequate account of justification. Because of the subjective character of the notion of an evaluation system, our account of personal justification must be combined with an objective constraint to yield knowledge. The problem immediately before us, however, is to analyze the notion of coherence in the schema above. We can execute this analysis by appeal to what it is reason-

able to accept on the basis of an evaluation system. We shall later provide an explanation of what makes it more reasonable to accept one thing than another on the basis of an evaluation system.

Let us begin by assuming that we are able to tell, at least intuitively, when it is more reasonable to accept one thing than another on the basis of our own evaluation system. Proceeding subjectively, I can dispense with certain skeptical objections in short order. Consider the skeptical objection that I might at this very moment be hallucinating, deceived by some powerful demon, or having my disembodied brain stimulated electronically by some scientist.[1] On the basis of my evaluation system, it is more reasonable for me to accept that I see a cat before me than that I am hallucinating. Why? I accept that I can tell when I am not hallucinating. I accept that I have refrained from ingesting hallucinogenic substances and that there is no indication in my experience of hallucination. I may not be able to say exactly how I can tell that I am not hallucinating at the moment, but I accept my trustworthiness in the matter. This is not, of course, a refutation of the skeptic. It is a statement of the consequences of what I accept.

Similarly, I accept that I can tell that I am not now deceived by a powerful demon or a powerful scientist, though, again, I may not be able to say how I can tell that these things are so. Suppose some skeptic suggests that everything I accept is the result of the brainwashing efforts of a Cartesian demon powerful enough to determine completely what I accept. With a twisted grin, the skeptic might note that I might be completely alone in the world with the exception of the undisclosed demon and not have a single external object on which to sit in my epistemic desperation.

Confronted by such a skeptic, I have two possibilities. I can accept what the skeptic says. In that case I shall be silenced, possibly becoming demented, and there is no cure for that in epistemology. Or I can reject what the skeptic says while admitting the logical consistency of the skeptical fantasy. I accept the trustworthiness of my senses and reason while admitting the logical consistency of the skeptical oddity. I prefer accepting the trustworthiness of senses and that there is no such demon to accepting that I am deceived by such a demon and, on the basis of this evaluation, it is more reasonable for me to accept the existence of others and of the objects of the external world than to accept the existence of the deceptive demon.

Suggestion of a Unified Theory of Knowledge

In general, how reasonable it is for me to accept something will depend on what I accept about my trustworthiness in the matter. When I have no reason to trust some source of information, then my evaluation system fails to provide any basis for considering it more reasonable to accept the informa-

tion I receive from that source than to reject it. On the other hand, when I consider myself and my source of information to be trustworthy, I can answer the skeptical objection on the basis of my evaluation system from what I accept. A skeptic denies that I perceive an external world. I answer that it is more reasonable for me to accept that my senses are worthy of my trust and do not systematically mislead me than to accept the opposite. A skeptic denies that my inductive reasonings lead me to truth. I answer that it is more reasonable for me to accept that my inductive reasonings are worthy of my trust, fallible though they be, and usually lead me to truth when they are judicious than to accept the opposite. A skeptic denies that my mathematical powers enable me to reason validly. I answer that it is more reasonable for me to accept that (though I sometimes err) my mathematical powers, when used with circumspection, allow me to reason validly than to accept the opposite.

In all this, I appeal to what I accept and do not expect the skeptic to concede what I say. We will consider the position of the skeptic carefully in Chapter 9 and appreciate her contribution. It is, nonetheless, important to note that if I appeal to my evaluation system and what I really do accept, I find that I have replies to skeptical objections that are genuine and based on my conception of myself, the world, and my relationship to it. They are sufficient, moreover, to suggest the possibility of a unified account of reasonableness, justification, and knowledge—one that combines the contributions of reason and experience.

These remarks, promissory as they are, reveal a connection between coherence with an evaluation system and the reasonableness of accepting something. Some claims conflict with others, as my claim, for example, that I see a cat conflicts with the skeptical claim that I am hallucinating. If it is more reasonable for me to accept one of these conflicting claims than the other on the basis of my evaluation system, then that claim fits better or coheres better with my evaluation system. The claim that I see a cat coheres with my evaluation system while the conflicting claim that I am hallucinating does not cohere with that system. My evaluation system does not supply me with the original conviction that I am seeing a cat; perception does that. On the other hand, my evaluation system adjudicates in favor of the conviction against those with which it conflicts and justifies me in accepting the conviction. Thus, what I accept coheres with my evaluation system if that system favors what I accept over objections against it, those of the skeptic, for example.

The foregoing reflections suggest the following preliminary definition:

> p coheres with the evaluation system of S at t if and only if it is more reasonable for S to accept p than to accept any *objection* to it on the basis of the evaluation system of S at t.

The Evaluation System

The notion of an evaluation system and of an objection against a target claim have been informally explained but should now be defined; first, we define the notion of an acceptance system:

> The acceptance system of *S* at *t* is by definition the set of states of acceptance of *S* described by statements of the form—*S* accepts that *p*—attributing to *S* just those things *S* accepts at *t* with the objective of *obtaining* truth and avoiding error with respect to the content accepted, that is, with respect to the content that *p*.

Next we define the notion of the system of preferences concerning or over acceptances, the preferences system of S, as follows:

> The preference system of *S* at *t* over acceptances is by definition the set of states of preferences described by statements of the form—*S* prefers accepting that *p* to accepting that *q*—attributing to *S* just those preferences *S* has at *t* with the objective of obtaining truth and avoiding error with respect to the content of the acceptances.

Then we define the notion of the system of reasoning concerning or over acceptances, the reasoning system of *S*, as follows:

> The reasoning system of *S* at *t* over acceptances is by definition the set of states of reasoning described by statements of the form—*S* reasons from acceptance of the premises p_1, p_2, and forth to p_j to acceptance of the conclusion *c*—with the objective of obtaining truth and avoiding error with respect to the content of the acceptances.

Finally, we define the notion of a system of evaluation, the evaluation system of *S* as follows:

> The evaluation system of *S* at *t* is by definition the combined set of states of the acceptance system, the preference system, and the reasoning system of *S* at *t*.

Objection Defined

The notion of an objection must also be defined, for it is also problematic. One might think that objections to a target claim are only those opposing claims that contradict it. Opposing claims that contradict a target claim are

objections to it, but some claims that do not contradict the target claim may also constitute objections to it. An objection to a target claim need not contradict the claim. The instance of hallucination was such an example. The claim that I see a real cat is not contradicted by the claim that I am hallucinating, for it is logically possible for a person to actually see a real cat even though he is hallucinating. The claim that I am hallucinating does not logically conflict with the claim that I see a real cat, but the assumption that I am hallucinating would make it less reasonable for me to accept that I am seeing a real cat than the opposite assumption. If I am hallucinating, then I am less trustworthy about what I see than if I am not hallucinating. The desired notion of an objection to a target claim may be defined in terms of a conception of reasonableness on an assumption as follows:

> *o* is an objection to *p* for *S* on system *X* at *t* if and only if it is less reasonable for *S* to accept that *p* on the assumption that *o* is true than on the assumption that *o* is false on the basis of the system *X* at *t*.

The foregoing definition contains a variable *X* that may be replaced by reference to the evaluation system of *S* to obtain the required notion of an objection on the basis of an evaluation system. The definition is given in this general form so that we may avail ourselves of the definition in order to define an objection on the basis of other systems later. The definition of an objection on the basis of acceptance system obtained from the foregoing is as follows:

> *o* is an objection to *p* for *S* on the basis of the evaluation system of *S* at *t* if and only if it is less reasonable for *S* to accept that *p* on the assumption that *o* is true than on the assumption that *o* is false on the basis of the evaluation system of *S* at *t*.

Having defined an objection in this way, we must now reconsider whether our definition of justification in terms of coherence is adequate. If it is more reasonable to accept something than to accept any objection to it, it is natural to think of the objections to *p* as having been answered. The notion of *an objection being answered* may be defined as follows:

> An objection *o* to *p* is answered for *S* on *X* at *t* if and only if *o* is an objection to *p* for *S* at *t* and it is more reasonable for *S* to accept that *p* than to accept that *o* on *X* at *t*.

These definitions suggest that a person is personally justified in accepting something just in case it coheres with the evaluation system of the person in

the sense of answering all objections on the basis of the evaluation system. Therefore, we might define personal justification as follows:

> *S* is personally justified in accepting that *p* at *t* if and only if every objection to *p* is answered on the basis of the evaluation system of *S* at *t*.

We shall attempt to clarify the implications of this definition with some examples and then note a defect that requires some amendment of it.

The Justification Game: Replying to a Critic

How are we to decide whether an objection to a target claim is answered? It is not necessary that a person has reflected on the objection for the objection to be answered, but it is necessary that the evaluation system of the person imply that it is more reasonable to accept the claim than the objection. If the evaluation system of a person implies that it is more reasonable to accept that *p* than to accept that *c,* then the person must be in a state to think and reason as though this were true. We can determine that this is so by imagining how a person would respond to critical questions. We imagine a game a person plays with a critic, who is a kind of internalized skeptic, to show that she is personally justified in accepting what she does. Let us refer to the game as the *justification game.*

The justification game is played in the following way. The claimant presents something she accepts as true. The critic may then raise any objection to what the claimant presents. If what the claimant accepts is something that is more reasonable for her to accept than the critical objection, that is, if the objection cited by the critic is answered, then the claimant wins the round. If all the objections raised by the critic are answered, then the claimant wins the game. If she wins the game, she is personally justified in accepting what she presented; if not, she is not personally justified. The game is a heuristic device for understanding the considerations that make a person justified in accepting something rather than a psychological model of mental processes. The justification game is one that the evaluation system of person permits her to play rather than one in which she has actually engaged.

Let us consider a few rounds of the justification game played by an imaginative critic with me as the claimant. In this game, the critic objects to a claim of mine. The claim may be considered as a statement of something I accept or, to give a little added vivacity to the game, to something I claim to know. In earlier chapters, we noted that our justification for what we ac-

cept depends on background information. This information is contained in the evaluation system and accounts for the reasonableness of the reply to the critic. It is what makes it more reasonable for me to accept what I do than the objections presented by the critic.

Imagine that I am at the Edinburgh Zoo looking at a zebra. The animal is a paradigm example of a zebra and the sign before me says 'zebra.' I would claim to know that I see a zebra. So I enter that as the claimant in the justification game.

Claimant: I see a zebra.
Critic: You are asleep and dreaming that you see a zebra.
Claimant: It is more reasonable for me to accept that I see a zebra than that I am asleep and dreaming that I see a zebra. (I can tell that I am awake and not asleep and dreaming now. My experience does not feel at all like a dream and I have a distinct memory of what preceded my present experience, leaving my hotel, taking the cab to the zoo, buying a ticket, all of which is trustworthy information that I am now at the zoo looking at a zebra and not asleep and dreaming.)
Critic: You are awake but hallucinating a zebra.
Claimant: It is more reasonable for me to accept that I see a zebra than that I am hallucinating a zebra. (There is nothing in my experience that would lead me to think that I am hallucinating, nor does my memory of the past indicate that there is any reason to think that I might be hallucinating now. I have not ingested any hallucinogenic substances, I am not deprived of sleep or food in any extreme manner, and in general I have no indication that I am in an abnormal state. There is no indication that I am hallucinating and there is trustworthy evidence that I am not.)
Critic: You are seeing a mule painted with stripes to look like a zebra.
Claimant: It is more reasonable for me to accept that I see a zebra than that I see a mule painted with stripes to look like a zebra. (Though I have no specific information about the stripes, I have no reason to believe that the Edinburgh Zoo would paint a mule to look like a zebra and identify it as a zebra, or be deceived by somebody else doing so. The Scots are known for their honesty. So my perceptual evidence is trustworthy information that the animal is a zebra.)
Critic: You are generally deceived in a systematic way and see nothing. You are either a disembodied mind deceived by a Cartesian demon or a disembodied brain lying in a vat deceived by electrical information supplied to the brain by a scientist.
Claimant: It is more reasonable for me to accept that I see a zebra than that I am generally deceived in a systematic way and see nothing. (I have no reason to think that I am deceived in a systematic way and, though it

would be impossible to detect such a systematic deception, the hypothesis is totally improbable on my evidence. The evidence that I see a zebra is trustworthy and renders it very improbable that I am deceived.)

The justification game played to the point reached above leaves the claimant the victor in each round. There is no pretension that the claimant has refuted the critic, for the claimant has appealed to his evaluation system and to what he accepts, which might not be admitted by a genuinely skeptical critic. The claimant's replies are, however, adequate replies for the purposes of exhibiting personal justification, for showing what justifies him on the basis of his evaluation system. The parenthetical remarks illustrate the sense in which coherence with my evaluation system is what personally justifies me in accepting what I do, that I see a zebra. Each challenge of the critic is rejected on the grounds that it does not cohere with what I accept, my preferences over what I accept, and my reasonings concerning what I accept, whereas my claim that I see a zebra does cohere with that system. The parenthetical remarks are part and parcel of my evaluation system. They are a brief summary of the relevant aspects of it.

Answering Objections

There is a problem for the theory of personal justification that is more difficult to solve and will require an amendment of our notion of personal justification. The problem is that some objections a critic might raise cannot be expected to be answered, at least in a sense defined, in cases in which a person is personally justified in accepting something and indeed knows that what he thus accepts is true. The reason is that some objections to a claim may be very indirect and may, as a result, be very reasonable to accept. Our definition of answering an objection requires that for an objection to be answered the target acceptance must be more reasonable to accept than the objection. A very reasonable objection might, as a result, be one that a person cannot answer in this sense. But we shall find cases in which the person is justified in accepting what she does and knows what she accepts is true even though there are very reasonable objections. They must be dealt with in another way.

Let us return to the justification game between myself as the claimant and the critic. Consider the zebra example again.

Claimant: I see a zebra.
Critic: People sometimes dream that they see zebras.

Suppose that I remember having had a very strange dream full of strange episodes, not at all like everyday life, in which I dreamed I saw a zebra. Though I have not the least doubt about whether I am dreaming at the moment—that is, I am sure that I am at the Edinburgh Zoo looking at a zebra—I must concede that the critic has raised an objection. Let us compare the following two assumptions:

A. People sometimes dream that they see zebras.
NA. People never dream that they see zebras.

It is more reasonable for me to accept that I see a zebra on the assumption that *(NA)* is true than on the assumption that *(A)* is true. Why? If people sometimes dream that they see zebras, which they do, then they might be misled into accepting that they see zebras when they are only dreaming that they do: whereas if they never dream that they see zebras, then they will not be misled in this way. If people are misled in this way, then, of course, I might be misled in this way.

It is clear what the reply should be, namely, that *(A)* is irrelevant because I am not dreaming. Thus, I should be allowed a step in the game which, though it does not consist of answering the objection offered by the critic, does neutralize the critic's objection. Thus, I should be allowed to proceed in the game as follows:

Claimant: I see a zebra.
Critic: People sometimes dream that they see zebras.
Claimant: I am not dreaming.

Under what conditions, though, should neutralizing replies like the last be allowed?

The answer depends on the reasonableness of accepting the neutralizer in conjunction with the objection. If it is as reasonable for me to accept the objection together with the neutralizer as to accept the objection alone, then I may use the neutralizer as a reply to the critic in the justification game. Compare the following two statements:

O. People sometimes dream that they see zebras.
O&N. People sometimes dream that they see zebras, but I am not dreaming.

On the basis of my acceptance system, it is as reasonable for me to accept the latter statement as to accept the former. There is, of course, some addi-

tional risk of error added to the latter, but the objective of accepting what is true supplements the objective of avoiding error. Put another way, *(O)* gives us less of the relevant truth about dreaming and seeing zebras than *(O&N)* does. The critical claim in the justification game outlined above is true, but it is misleading in the context because it suggests that I might be dreaming that I see a zebra. The neutralizing claim that I am not dreaming corrects this misleading suggestion. It successfully neutralizes the critical move. The conjunction *(O&N)* is, therefore, as reasonable for me to accept as *(CO)* alone. We may define the notion of neutralization as follows:

> *n* neutralizes *o* as an objection of *p* for *S* on *X* at *t* if and only if *o* is an objection to *p* for *S* on *X* at *t*, but the conjunction of *o* and *n* is not an objection to *p* for *S* on *X* at *t*, and it is as reasonable for *S* to accept the conjunction of *o* and *n* as to accept *o* alone on *X* at *t*.

The justification game should be amended to allow that the claimant be in a position either to beat or neutralize an objection introduced as a move by the critic. An objection may be neutralized by conjoining a neutralizer to the objection and noting that the conjunction is as reasonable to accept as the objection.

The Justification Game and the Definition of Personal Justification

Reformulating the justification game to conform to this format, it would run as follows:

Claimant: I see a zebra.
Critic: People sometimes dream that they see zebras.
Claimant: It is as reasonable for me to accept both that people sometimes dream that they see zebras *and* that I am not dreaming as to accept the former alone. (My information that I am not dreaming is trustworthy and the information that people sometimes dream that they see zebras is misleading with respect to the question of whether I see a zebra. Conjoining that I am not dreaming produces a result that is not an objection to my claim that I see a zebra.)

A claimant wins a round in the justification game just in case she can answer or neutralize the objection produced by the critic. A claimant wins the justification game by showing that she is justified in accepting that *p*, just in case she can win every round in the justification game on the basis of her

evaluation system concerning the claim that p. We may put this more formally in terms of a general definition of justification as follows:

> S is justified in accepting that p at t if and only if everything that is an objection to p for S on X at t is either answered or neutralized for S on X at t.

Personal justification based on the acceptance system of a person may then be defined as follows:

> S is personally justified in accepting that p at t if and only if everything that is an objection to p for S on the basis of the acceptance system of S at t is answered or neutralized on the basis of the acceptance system of S at t.

This completes our definition of personal justification.

A Foundationalist Objection

One traditional foundationalist objection concerning our account of personal justification remains to be met. A detractor will point out that the remarks made in the justification game in reply to the critic might be unjustified. How, she might inquire, can I be justified in accepting that I see a zebra on the basis of things that I accept but am not justified in accepting? The fundamental reply is that each of the things that I accept may be claims that I am justified in accepting because of other information I accept. Still, the foundationalist might persist, suppose that you only accepted one thing, then it would cohere with itself, and surely that is not the sort of justification required for knowledge. Moreover, even if a person accepts many things, it is possible that one claim stands in isolation from everything else. Again, since that one thing coheres with itself, the person will, unfortunately, turn out to be personally justified in accepting that one claim as a simple result of accepting it.

How are we to reply to the foundationalist? First of all, mere acceptance of something by itself does not suffice to yield the result that it is more reasonable to accept it than the objections to it, and the foundationalist objection fails for that reason. If mere acceptance does not suffice, what does? From the short justification game played above, the answer should be apparent. It is not enough that one accept something for it to be more reasonable than the objections to it on the basis of one's evaluation system. One must have some information that such acceptance is a trustworthy guide to

truth. The objective of acceptance is to obtain truth and avoid error in the specific thing accepted. For it to be reasonable to think one has succeeded, one must have information to meet the objections of a critic in the justification game. If I put forth the claim that I see a zebra and the critic counters with the claim that I do not, it is no answer to the critic to say that I accept that I see a zebra. I must have reason to think that I can tell a zebra when I see one in circumstances like those I am in at the moment and, consequently, that I am trustworthy in such matters.

The Principle of Trustworthiness of Acceptance

The foregoing answer leads to the final dialectical move. The foundationalist will surely note that everything now depends on the claim that my acceptance is a trustworthy guide to truth and that I am trustworthy, as I aver. She will inquire how that claim is itself justified. The claim that I am trustworthy in any particular matter under any special set of circumstances may be justified on the basis of the other things that I accept; I accept that I have had success in reaching the truth about similar matters in similar circumstances in the past and that the present circumstances do not differ in any relevant way from past circumstances when I was correct. There is, however, more to the issue. I may accept that my faculties, perception, memory, reasoning, and so forth are trustworthy guides to truth in circumstances of the sort that I find myself in when I accept what I do. I must accept, moreover, that I am worthy of my own trust, that is, that *I* am trustworthy as well: that when I accept something, that is a good enough reason for thinking it to be true, so that it is reasonable for me to accept it.

Thus, there is one special principle of an evaluation system, to wit, that one is trustworthy (worthy of one's own trust) in how one seeks to obtain truth and avoid error in what one accepts to these ends. This amounts to the following principle formulated in the first person:

> *T.* I am trustworthy (worthy of my own trust) in what I accept with the objective of accepting something just in case it is true.

If someone else accepts that I am trustworthy in this way, then my accepting something will be a reason for her to accept it. Similarly, if I accept that I am trustworthy in this way, then my accepting something will be a reason for me to accept it. Another person might be confronted with some other considerations that cast enough doubt on whether what I accept is true, even granting my trustworthiness, so that my accepting something, though providing a reason for her accepting it, does not always justify her

in accepting what I do. My accepting something when I am confronted with similar considerations casting doubt on whether what I accept is true would fail to justify me in accepting it as well.

Reasonableness and Trustworthiness

A consequence of adding principle *(T)* to my evaluation system is that I may reason from it and the acceptance of some target acceptance that *p* to the conclusion that the target acceptance is reasonable. My reasoning would be as follows:

T. I am trustworthy (worthy of my own trust) in what I accept with the objective of accepting something just in case it is true.
I accept that *p* with the objective of accepting that *p* just in case it is true.
Therefore, I am trustworthy in accepting that *p* with the objective of accepting that *p* just in case it is true
Therefore, I am reasonable in accepting that *p* with the objective of accepting that *p* just in case it is true.

The argument from trustworthiness to reasonableness, which I shall refer to as the *trustworthiness argument*, assumes that my trustworthiness may explain why it is reasonable for me to accept what I do. I am reasonable in accepting what I do in order to obtain truth and avoid error because I am trustworthy in accepting what I do to obtain truth and avoid error. My reasonableness is explained in the argument by my trustworthiness. However, it is important to note that for the explanation to succeed, the premise of trustworthiness must be true. A false premise fails to explain anything.

The trustworthiness argument requires some qualification. First, the principle *(T)* must not be construed as a universal statement to the effect that I am always trustworthy in whatever I accept. It is a statement of a capacity and disposition to be trustworthy but, however capable and disposed I am to be trustworthy, I shall nonetheless fail now and then to be trustworthy in what I accept. Our capacities are fallible ones. Thus, the inference *(T)* to the conclusion that I am trustworthy in the target acceptance, the acceptance that *p*, is inductive rather than deductive. It is like the inference from the premise that my lawyer is trustworthy to the conclusion that he is trustworthy in the way he has constructed my will or from the premise that a city water supply is trustworthy to the conclusion that the water supplied in my glass is trustworthy.

Second, my trustworthiness in what I accept is not simply a matter of my current rate of success in obtaining truth and avoiding error in what I accept. I may proceed in a manner that is worthy of my trust in what I accept but be deceived through no fault of my own. Suppose that I seek to obtain

truth and avoid error with the greatest intellectual integrity, but I am deceived by some being more powerful than myself, the evil demon of Descartes, perhaps. Then I am worthy of my trust in what I accept though I am deceived. I am as trustworthy as the circumstances allow.

Moreover, even in perfectly normal circumstances where no one seeks to deceive me, I may proceed in a manner worthy of my trust without present success in my quest for truth. My trustworthiness is not just a matter of what I now accept, but also of how I change what I accept and even of how I change my methods of changing in order to correct what I accept and improve in my quest to obtain truth and avoid error. Thus, my present trustworthiness projects beyond the static moment dynamically into the future. I am trustworthy in what I now accept in part because I am trustworthy in how I change.

My trustworthiness is a consequence of how I learn from experience and learn from others, how I evaluate experience and how I evaluate others. It is even more fundamentally a matter of my readiness to consider objections to what I accept and either find satisfactory replies, thus succeeding in answering or neutralizing objections, or changing what I accept when I cannot deal with the objections in a trustworthy way. My trustworthiness is not an intellectual abstraction. It rests on a dynamic process of evaluation and amalgamation of information I receive from others and from my own experience. My present trustworthiness is at once personal and social, synchronic and diachronic, critical and revisionary.

The Trustworthiness of Reasoning

It becomes obvious, moreover, that the reasoning in the trustworthiness argument must also be trustworthy and presupposes a principle concerning the trustworthiness of reasoning corresponding to principle *(T)* as follows:

> *(R)* I am trustworthy in how I reason with the objective of concluding something just in case it is true.

Starting with this principle I may reason to the conclusion that I am reasonable in accepting the conclusions of reasonings, including the reasoning of the trustworthiness argument itself. The reasoning is as follows:

> I reason to the conclusion that *c* with the objective of concluding *c* just in case it is true.
> *Therefore,* I am trustworthy in my reasoning to the conclusion that *c* with objective of concluding *c* just in case it is true.
> *Therefore,* I am reasonable in my reasoning to the conclusion that *c* with the objective of concluding *c* just is case it is true.

As in the case of the trustworthiness argument, principle *(R)* must be read as describing a disposition to be trustworthy and the steps of the argument must, as a result, be considered inductive rather than deductive. The last step of the argument is based on the assumption that my trustworthiness in how I reason explains my reasonableness in how I reason. Assuming, then, the reasonableness of the conclusions of my reasoning, the conclusion of the trustworthiness argument, that I am reasonable in accepting that p, is itself reasonable.

The Trustworthiness of Preference

Having accepted the trustworthiness of acceptance and reasoning, it is clear that I should add the trustworthiness of preferences concerning acceptance formulated as follows:

(P) I am trustworthy in what I accept with the objective of preferring to accept something just in case it is true.

It would be repetitious to repeat the argument for the reasonableness of preference based on the trustworthiness of preference. I would proceed by reasoning from the premise that I am trustworthy in what I prefer accepting in the quest for truth to the conclusion that I am reasonable in what I prefer accepting in that quest. The reasonableness of preference is a consequence of the reasonableness of accepting the premise and reasoning to the conclusion.

However, it is important to notice the special role of the conclusion concerning the reasonableness of preference. Our theory of justification rests on the possibility of evaluating whether it is more reasonable to accept a target acceptance than an objection to it. How can we account for the reasonableness of preference? It is explained by the trustworthiness of preference and the resulting reasonableness of preference. If I prefer accepting a to accepting o and my preference is reasonable, then it is more reasonable for me to accept a than to accept o. So we have an account of why it is more reasonable to accept one thing than another based on preference. Of course, the trustworthiness and reasonableness of preference may in turn be explained by the trustworthiness and reasonableness of acceptance and reasoning. Once we have explained the reasonableness of preference, the comparative reasonableness of a target acceptance over objections to it is an immediate result of preference for accepting one over the other. Trustworthiness explains the reasonableness of preference. The reasonableness of preference explains reasonableness of comparative evaluation on the basis of the evaluation system.

The manner in which we trust what we accept, what we prefer, and how we reason indicates that we do accept that we are trustworthy. The mark of

our regarding a person as trustworthy is that we trust him or her, and this applies to ourselves as well. Some more restricted principles of acceptance may supplement the unqualified one in a reflective person, forcing her to arrive at the conclusion that she is less trustworthy in some domains than others. For example, despite her best efforts not to accept things without adequate evidence, she notes how often she commits a kind of doxastic *akrasia* and accepts some things without adequate reason even though she thinks she ought not to do so. For example, she might be attracted to particularly elegant mathematical principles and, as a result, accept some principles as theorems because of their elegance, without adequate consideration of the proofs offered for them. Nevertheless, she seeks, if she is trustworthy in her quest for truth, to correct her ways and improve upon how she proceeds in her quest for truth. This striving for improvement and the changes that result from it accounts for her present trustworthiness and sustains principle *(T)* even given her fallibility, which she recognizes and we all share.

The Virtuous Loop of Reason

A person may appeal to principle *(T)* and the trustworthiness argument to defend the reasonableness of accepting what she does, but what defense should she offer in favor of *(T)* itself? She may, of course, appeal to the character of what she accepts, to the various things she accepts, and reason inductively from premises concerning the trustworthiness of individual acceptances in support of the conclusion that *(T)*. She might reflect on what she has accepted and her fine track record of mostly accepting what was worthy of her trust to accept. This argument would establish that the trustworthiness of her acceptances manifests her disposition to be trustworthy in what she accepts. There is, moreover, another argument, a more direct one, for the reasonableness of accepting *(T)* that she may employ. If a person accepts *(T)*, then her acceptance of *(T)* itself will have the result that it is reasonable for her to accept *(T)* by an application of the trustworthiness argument to *(T)* itself as the target acceptance p. The principle applies to itself. It yields the results that if she accepts *(T)* with the objective of accepting it just in case it is true, then she is trustworthy in accepting it, and by the trustworthiness argument, to the conclusion that she is reasonable in accepting it. Once again, however, principle *(T)* must be true for the argument to establish the conclusion. If the person is demented and is not trustworthy, no argument the person can use will succeed in proving that she is trustworthy or reasonable in what she accepts.

It is, moreover, natural simply to regard a person as applying principle *(T)* to itself. As such, it can play a special role. The addition of the principle *(T)* to an evaluation system has the result when combined with principle

(R) that other things we accept are reasonable for us to accept. It has the additional consequences that it is reasonable for us to accept *(T)* itself. To borrow an analogy from Thomas Reid, just as light, in revealing the illuminated object, at the same time reveals itself, so the principle, in rendering the acceptance of other things reasonable, at the same time renders the acceptance of itself reasonable.[2]

This does not entail, as the foundationalist might wish, that we are personally justified in accepting the principle, only that it is reasonable to accept it. Recall that it might be reasonable for somebody to accept something she is not justified in accepting because justification requires something beyond reasonableness, namely, that all objections to the target claim can be met. Some objection to principle *(T)* might not be answered even though it is reasonable to accept the principle. One such objection is the fallibilistic claim that we are sometimes in error in what we accept. To meet such an objection, we need more information about the sort of circumstances in which we err and those in which we do not. Thus, even in the case of principle *(T)*, we require some background information in order to be personally justified in accepting the principle. If, however, the foundationalists are incorrect in arguing that there are basic beliefs that justify themselves, they are right in thinking that there are some beliefs that may contribute along with other beliefs to their own justification. For *(T)*, though supported by other acceptances, provides a premise for concluding that it is reasonable to accept *(T)* itself.

Should we conclude that *(T)* is a basic belief? That would be incorrect, since *(T)* depends for its justification on the background system of other things we accept, including *(R)*. The better metaphor would be that of a keystone in an arch. The keystone is a triangular stone inserted in the top of an arch. It supports the arch, for the arch would collapse were it removed; at the same time, it is, of course, supported by the other keystones in the arch. We may think of the stones in the arch as the acceptances in the acceptance system and the principle *(T)* as the keystone.

There is obviously a circularity in the trustworthiness argument when we use the principle *(T)* as a premise to support the conclusion that the other acceptances are reasonable and then use those acceptances and the principle itself to conclude that it is reasonable to accept it. Should we find the circle vicious? To use a premise to prove something to a skeptic who challenges it violates the rules of rhetoric. But to explain why it is reasonable to accept what we do, the circle may be virtuous. If we have a principle that explains why it is reasonable to accept what we do, it is a virtue rather than a vice that it should at the same time explain why it is reasonable to accept the principle itself. The other alternative is that the principle should be a kind of unexplained explainer that explains why it is reasonable for us to accept the other things we accept and then falls mysteriously silent when asked why it

is reasonable to accept the principle itself. The loop by which the principle explains why it is reasonable to accept the principle as well as other things is worthy of epistemic praise rather than rhetorical condemnation.

Reasonableness and Probability

Coherence and personal justification have been defined in terms of reasonableness and acceptance directed at obtaining truth and avoiding error. Reasonableness has been taken as undefined, though explained in terms of trustworthiness. This has two constructive advantages. First, we acknowledge the normative aspect of justification, which tells us what we ought or ought not accept to achieve our purposes. One ought not accept something, if one is epistemically rational, when it is more reasonable not to accept it. At the same time, we leave open the question of whether this notion of reasonableness is reducible to some naturalistic or nonnormative conception. Second, we allow for a plurality of factors to influence the normative evaluation. We have left it open what considerations might make it more reasonable to accept one thing rather than another in the quest for truth on the basis of one's evaluation system. We leave open the question of what sort of factors are relevant to obtaining truth and avoiding error. Most of the defenders of the coherence theory—Quine, Sellars, Harman, Aune, Lycan, Rosenberg, and Bonjour—have in one way or another proposed that multiple factors determine whether a belief coheres with some system.[3] They have, however, differed among themselves as to what factors are relevant.

Is some naturalistic reduction of reasonableness possible? The simplest reduction would be to equate reasonableness with probability. For such an account to be naturalistic, we would have to be sure that the notion of probability was itself free from normative definition. This condition is not satisfied in those notions of probability which impose normative constraints on the assignment of probabilities. The equation of reasonableness with probability fails for other reasons, however. One needs to consider more than probability to decide whether it is more reasonable for a person to accept one thing than another. Since the equation of reasonableness with probability has so much intuitive lure and traditional backing, it is worth reconsidering briefly why it should be rejected.

The basic reason for rejecting the equation of probability and reasonableness is that probability is only one factor relevant to deciding what is reasonable to accept in the interests of obtaining truth and avoiding error. One can see this from a simple example. Compare the following two claims:

It looks to me as though there is a computer in front of me.
There is a computer in front of me.

How would one compare the reasonableness of accepting each of these statements with the objective of obtaining truth and avoiding error? The first statement is less risky, but it tells us less. The second statement is a bit more risky, but it is more informative. As a result, we can say that the risk of error is greater in accepting the second than the first, but the gain in obtaining truth is greater in accepting the second than the first. This example reveals that the objectives of obtaining truth and avoiding error are distinct and may pull in opposite directions. The more informative a statement is, the more it tells us about the world, the greater our gain in accepting it if it is true, and the greater our risk of error. The probability of a statement tells us what our risk of error is, but it tells us nothing about how much we gain in accepting it when it is true.

Major scientific claims, those concerning galaxies, genes, and electrons, for example, though among the most important things we accept and claim to know, are less probable than either of the cautious claims articulated above. The reasonableness of accepting such claims is influenced by our interest in accepting claims which, if true, are important general truths about ourselves and our universe. Given the past history of scientific claims, even those put forth by scientists of great genius, we must concede that the risk of error, the probability that these claims are false, is far from negligible. We can easily see that the interest in accepting what is true and the interest in avoiding accepting what is false may pull in opposite directions by considering the results of aiming at one to the exclusion of the other. If a person were only interested in avoiding error and indifferent to accepting truths, total success could be attained by accepting nothing at all. If, on the other hand, a person were only interested in accepting everything that is true and indifferent to accepting falsehoods, total success could be attained by accepting everything. The problem is to accept what is true while at the same time seeking to avoid error.

Reasonableness and Expected Utility

The foregoing ideas can be summarized in a simple mathematical representation. Suppose that we could specify what value or, more technically, the positive utility we assign to accepting some specific hypothesis, *h*, if *h* is true, and represent that by '*Ut(h)*' and, similarly, the negative utility we assign to accepting *h*, if *h* is false, and represent that by '*Uf(h).*' If we ask ourselves how reasonable it is to accept *h* in the interests of accepting what is true and avoiding accepting what is false, we must take into account the values of both *Ut(h)* and *Uf(h)*. There are just two outcomes of accepting *h*, that *h* is true and that *h* is false, and we must take into account what value we attach to each of those outcomes. There is, however, another factor to consider, namely, how probable each of those outcomes is. So, if we let '*r(h)*' represent the degree of reasonableness of accepting *h*, and '*p(h)*' rep-

resent the probability of *h* being true and *'p(-h)'* as the probability of *h* being false, we obtain the following formula for computing the reasonableness of accepting *h:*

$$r(h) = p(h)Ut(h) + p(-h)Uf(h).^4$$

The reasonableness of accepting *h, r(h),* the sum of two products, may be called the epistemic expected utility of accepting *h.* It is equal to the sum of one's positive expectation of accepting *h* when *h* is true and one's negative expectation of accepting *h* when *h* is false.

This formula clarifies precisely the consequences of identifying probability and reasonableness. To do so is mathematically equivalent to assigning a value of one to *Ut(h)* and a value of zero to *Uf(h),* no matter what *h* is. In that case, no matter what claim *h* might be, $r(h) = p(h)$. It is absurd, however, to regard all truths as equally worth accepting. Some truths are more valuable than others measured strictly in terms of obtaining truth about the world because, as already noted, some truths tell us much more about the world than others. Therefore, the equation of reasonableness and probability must be rejected. Its rejection does not, of course, mean that probability is irrelevant to reasonableness. On the contrary, we can see that if the utilities of accepting two claims are the same, then the comparative reasonableness of accepting one in comparison to the other will be determined solely by the probabilities. This is more important than it might at first seem because some competing claims satisfy the constraint of having the same utilities.

The Lottery Paradox

To see the importance of the preceding observation, consider Henry Kyburg's lottery paradox.[5] The paradox proceeds from the assumption that some probability less than unity is sufficient for justified acceptance. Suppose that we pick a probability of .99 as sufficient. Consider, then, a lottery we know to be fair with one hundred tickets such that the winner has been drawn. In that case, any one of us could argue in the following manner: I am justified in accepting that the ticket number one has not won because the probability of its winning is only .01 and, therefore, the probability of its not winning is .99 as required. By the same argument, I am justified in accepting that each ticket has not won, for the probability of each ticket winning is .01 and of not winning is .99. The assumption that .99 is sufficient for justified acceptance will not permit me to argue that the conjunction of all these individual conclusions is justified, but it will enable me to argue that I am justified in accepting each member of a set of conclusions of the form, ticket *X* has not won. The set of conclusions I am justified in ac-

cepting to the effect that each of the X tickets has not won is, however, logically inconsistent with my knowledge that one of them has won.

Though it may be improbable that a given ticket has won, there is a simple argument to the effect that I do not know that it has not won and hence that I would not be justified in claiming to know, or accepting, that it has not won. I know that exactly the same reasoning is available to 'justify' accepting that the winning ticket has not won. The definition of justification given above, when combined with the formula for reasonable acceptance, yields the correct result that I am not justified in accepting that the number one ticket has not won. Consider the following move in the justification game:

Claimant: The number one ticket has not won.
Critic: The number two ticket has not won.

The critic has produced an objection to my claim because, by definition, o is an objection to p just in case it is more reasonable to accept that p on the assumption that o is false than on the assumption that o is true. If what she has claimed is false and the number two ticket has won, then my claim must be true. On the other hand, on the assumption that what she has claimed is true, the probability of my claim is reduced to 98/99 because the number of potential winners is reduced to 99. In this case, the utilities of accepting the two claims, mine and the critic's, are obviously the same, and therefore the comparative reasonableness of the two claims is the same. Consequently, the critic's claim is not answered—it is as reasonable as mine—and it cannot be neutralized either.

The Advantages of Truth

We have noted that the utility of accepting h when h is true depends on how much h tells us, that is, on how informative h is. Therefore the reasonableness or expected utility of accepting h is a function of the informativeness of h as well as of the probability of h. This observation sustains our earlier contention in the justification game that it may be as reasonable to accept some conjunction, that people sometimes dream they see zebras *and* I am not dreaming, as to accept only the first critical conjunct, that people sometimes dream they see zebras. The conjunction is less probable than the one conjunct because there is greater risk of error, but the conjunction is more informative and, given that the risk of error is negligible in either case, it is just as reasonable to accept the conjunction as the single conjunct.

This illustrates that it can be of greater advantage to accept one truth than another, depending on the characteristics of the truth, on its informativeness, for example. Other philosophers have insisted on other advan-

tages of accepting a truth. They have insisted on the advantages of explanatory power, of simplicity, of pragmatic value, and even of conserving what one has already accepted, as noted in the last chapter. Any of these factors may be relevant to the utility and expected utility of accepting something and, therefore, to the reasonableness of acceptance. It is, however, the truth of what is accepted that is paramount. Consequently, there is a barrier of risk below which we should not fall. No matter how much explaining we may accomplish by accepting *h*, no matter how simple or informative *h* may be, we cannot reasonably expect to gain anything in our attempt to obtain truth and avoid error when the risk of error is too great. Accepting a false explanation explains nothing. Accepting a simple hypothesis that is false may be nothing but an error of oversimplification. Accepting something of great informational content when it is false is only to accept a great amount of misinformation. The advantages of conserving error are minimal. Though we may value the other advantages of accepting a truth, it is the truth of what we accept that produces those advantages. Spices may enhance the flavor of good ingredients, but if the ingredients are spoiled, enhancing the flavor increases the risk of our consuming food that is dangerous to our health. Explanation, simplicity, and informativeness are but the spices of truth.

Introduction to the Literature

The most important recent books defending the sort of view contained in this chapter are Laurence BonJour's *The Structure of Empirical Knowledge* and the precursor of the present book, Keith Lehrer's *Knowledge*. The theories of Lehrer and BonJour are discussed by a number of authors with replies by Lehrer and BonJour in *The Current State of the Coherence Theory*, by John W. Bender. For another article critical of Lehrer's theory, see John W. Bender's "Knowledge, Justification, and Lehrer's Theory of Coherence." Also see Nicholas Rescher, *The Coherence Theory of Truth*, for another version of the coherence theory. The application of decision theory to epistemic problems was developed by Carl G. Hempel in "Deductive-Nomological vs. Statistical Explanation" and by Isaac Levi in *Gambling with Truth*, both difficult but readable works for those unafraid of symbols.

Notes

1. Descartes *Meditations*, II; Hilary Putnam, "Brains in a Vat," in *Reason, Truth, and History* (Cambridge: Cambridge University Press, 1984).

2. Thomas Reid, *Essays on the Active Powers of Man*, from *The Works of Thomas Reid, D.D.*, ed. Sir William Hamilton (Edinburgh: James Thin, 1895), 617.

3. See W. V. O. Quine, *Word and Object* (Cambridge: MIT Press, 1960); Wilfrid Sellars, "Some Reflections on Language Games," in *Science, Perception, and Reality* (London: Routledge & Kegan Paul, 1963), 321–58; Gilbert Harman, *Thought* (Princeton: Princeton University Press, 1973); Bruce Aune, *Knowledge, Mind, and Nature* (New York: Random House, 1967); William Lycan, *Judgment and Justification* (New York: Cambridge University Press, 1988); Jay Rosenberg, *One World and Our Knowledge of It* (Dordrecht: Reidel, 1980); Laurence Bonjour, *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press, 1985).

4. Decision theory is applied to epistemic issues by Carl G. Hempel in his article, "Deductive-Nomological vs. Statistical Explanation," in *Minnesota Studies in the Philosophy of Science*, vol. 3, ed. Herbert Feigl and Grover Maxwell (Minneapolis: University of Minnesota Press, 1962), 98–169; by J. Hintikka and J. Pietarinen in "Semantic Information and Inductive Logic," in *Aspects of Inductive Logic* (Amsterdam: North-Holland, 1966); and by Issac Levi in *Gambling with Truth: An Essay on Induction and the Aims of Science* (New York: Knopf, 1967).

5. See Henry Kyburg, *Probability and the Logic of Rational Belief* (Middletown, Conn.: Wesleyan University Press, 1961), 167.

This page intentionally left blank

7

Coherence, Truth, and Undefeated Justification

SHOULD WE SAY that a person knows every acceptance to be true that she is personally justified in accepting provided only that it is also true? Having reflected on the importance of truth, we must confront the objection that the evaluation system providing personal justification for an acceptance could be mostly in error even though the acceptance itself was true. In this chapter, we shall construct a theory of how personal justification may be converted into undefeated justification that does not depend on error. To accomplish this, we need only amend the justification game introduced in the last chapter to allow the critic greater advantage. Our account of undefeated justification arises from a technical problem, but the solution has more general implications. It will, as we shall see at the end of this chapter, allow us to reduce knowledge to undefeated justification. Finally, our analysis has the consequence, which we shall consider in subsequent chapters, that skepticism is in error.

The Uncharitable Possibility of Error

Some authors, most notably Donald Davidson, deny the possibility that what we accept could be mostly in error.[1] That claim, if correct, would simplify our task, and so we begin by considering his argument. Davidson claims that a principle of charity in interpreting the beliefs of others requires that we interpret them in such a way as to make them turn out to be mostly true. Since charity begins at home, we are committed to the supposi-

tion that our beliefs, at least what we accept in the quest for truth, are mostly true. Assuming the principle of charity as a principle of interpreting the content of our beliefs, we must suppose that our beliefs are mostly true and that the problem of acceptance systems, in which most of what is accepted is false, would not arise. Thus, Davidson's proposed principle of charity is congenial to the account of justification offered here.

Unfortunately, the principle does not itself appear to be true. We may be in a position in which it would be uncharitable or at least doxastically imperialistic to interpret the beliefs of another in such a way that they are mostly true. Suppose I am a nominalist. I confront someone who is such a devout Platonist that he is careful to couch the content of every belief of his in Platonistic terms and to reject everything not couched in such terms. When I believe that water is wet, he believes that water exemplifies the universal wetness and, indeed, refuses to accept the simple claim that water is wet on the grounds that this is a de-Platonized and hence erroneous description of the Platonistic fact of exemplification. We have a similar difference of opinion about all matters of fact. When I believe that x is F, he believes that x exemplifies *Fness* and refuses to accept the simple claim that x is F on grounds of the Platonistic inadequacy of the description. He tells me, moreover, why he restricts his beliefs in this way. It is because of his devout commitment to Platonism. On the basis of what I have thus discovered about his Platonism, I would conclude from the perspective of a nominalist that he has succeeded in formulating all his categorical beliefs about the world in such a way that they are false. They all imply that something exemplifies some universal when, nominalism being true, nothing exemplifies a universal. Error dominates in his acceptance system.

Application of the principle of charity would require that I interpret as many of his beliefs as possible as true. To conform to this principle, I would need to interpret his beliefs in a de-Platonized manner. I would have to interpret him as a nominalist. It is not clear how I should proceed to do this, and such an interpretation would be absurd in any case. Such examples show that the principle of charity is at best a defeasible method for interpreting the beliefs of others and, by extrapolation, our own beliefs as well. There is, unfortunately, no conceptual absurdity or necessary falsehood involved in supposing that most of what a person believes is false. Thus, any person, like our Platonist, may turn out to have a prodigiously large collection of false beliefs. Personal justification does not automatically convert to undefeated justification as a result of the necessity of interpreting most of the acceptance system of a person as true, for there is no such necessity.

Personal justification is the basis of undefeated justification. We shall define undefeated justification shortly, but it is useful to consider the intuitive idea. Undefeated justification is justification that cannot be refuted by appeal to any error on the part of the subject who has the justification. Thus,

a person has a justification that is undefeated for accepting something exactly when the person is irrefutably justified. To say that a person is irrefutably justified is not to say that he is invulnerable to counterargument. It is, instead, to say that no counterargument based on rejecting something the person has erroneously accepted will succeed to refute his justification.

No one can be irrefutably justified in accepting anything that he is not personally justified in accepting, but someone can fail to be irrefutably justified in accepting something that he is personally justified in accepting. Anyone familiar with those who accept astrology as the basis of predicting the future or who accept the claim that the universe was created by God a few thousand years ago can illustrate the point. The latter sort of person may provide a reinterpretation of the data concerning the age of the universe by claiming that when God created the universe a few thousand years ago it was created in such a way as to provide evidence of a much more ancient existence, perhaps as a test of faith. An opponent might well need to concede that such a person is personally justified in accepting that the Alps have only existed for a few thousand years, but such an opponent, particularly an opponent from geology, would be disinclined to concede that his adversary is irrefutably justified in accepting that the Alps are of such recent origin.

Moreover, the reason for denying that the fundamentalist is irrefutably justified cannot be that he or she is ignoring the evidence. On the contrary, we might suppose that the fundamentalist is also a geologist, one who takes special pleasure in observing that God has created the world in such a way that even the most precise scientific examination of geologists will not provide a clue as to the true origin of the universe. That is revealed through faith, not science. If such a person is not irrefutably justified in accepting that the Alps are only a few thousand years old, this is not due to any lack of scientific acumen but to the falsity of a basic assumption and the consequences thereof. It is lack of truth, not lack of science, to which a critic must appeal.

Undefeated Justification

The foregoing suggests that undefeated justification is personal justification that is not based on error. If we agree that a person is personally justified in accepting something and also that the independent information cited in the justification game is correct information, then we should also agree that she is irrefutably justified. This suggests a technical notion of justification based on what remains of the acceptance system of a person when all error is deleted. Let us call it *ultra justification.*

Ultra justification will require justification based on a new system that retains only what is true in the person's evaluation system or, to put it an-

other way, on what remains when everything false is eliminated from the person's evaluation system. Thus, acceptances of anything false are eliminated. Preferences for accepting something false over something true are eliminated. Reasonings that are unsound are eliminated. Nevertheless, the existence of the eliminated states of acceptance, preference, and reasoning in the original evaluation system must be acknowledged in the new system, for it is true they were in the original evaluation system. Let us call this new system, the basis for ultra justification, the *ultrasystem* for the person.

We may say the justification of a person is undefeated, or, what is the same thing, that the person is irrefutably justified, just in case the person is justified on the basis of the ultrasystem of that person. Thus, we may define justification that is undefeated or, equivalently, irrefutable justification, as follows:

> *S* is irrefutably justified (*S*'s justification is undefeated) in accepting that *p* if and only if *S* is justified in accepting that *p* at *t* on the basis of the ultrasystem of *S* at *t*.

The Ultra Justification Game and the Gettier Problem

We may illustrate the role of ultra justification by considering the justification game amended to produce the ultra justification game. The objections of the critic in the ultra justification game must be based on the ultrasystem of a person.

To clarify the role of the ultrasystem in irrefutable or undefeated justification, we may again consider a justification game, the ultra justification game, between the claimant and a new critic, whom we shall call the *ultracritic*. Suppose that we supply the ultracritic with a list that includes everything accepted by the claimant marked as to the truth or falsity of the thing accepted. The ultracritic is then allowed a new sort of move in the justification game. She may require that the claimant eliminate anything the claimant accepts that is false and, in addition, that he eliminate any preference concerning acceptance when the claimant prefers accepting something false to accepting something true and any reasoning involving the acceptance of false premises.

The ultracritic may then raise an objection that must be answered or neutralized after the claimant makes the required alterations, that is, eliminations, within his evaluation system, leaving him with his ultrasystem. The justification of the claimant is undefeated by any false statement just in case he wins this justification game, the ultra justification game, against the ultracritic.

As an example of the ultra justification game, consider the kind of counterexample Edmund Gettier raised against the claim that knowledge is justified true belief.[2] The claim is that someone in my class owns a Ferrari. Someone I know to be in my class, Mr. Nogot, says to me that he owns a Ferrari, shows me papers stating that he owns a Ferrari, and drives a Ferrari. I conclude that Mr. Nogot owns a Ferrari and, therefore, that someone in my class owns a Ferrari. In fact, another student in my class, Mr. Havit, owns a Ferrari, though I am entirely ignorant of this. The student I take to own a Ferrari, Mr. Nogot, does not own a Ferrari. He has lied to me, forged the papers, and so forth. Consider the following justification game:

Claimant: Someone in my class owns a Ferrari.
Ultracritic: None of the students in your class who appear to own Ferraris actually own Ferraris.
Claimant: It is more reasonable for me to accept that at least one student in my class, Mr. Nogot, owns a Ferrari than to accept that none of the students in my class own Ferraris. Mr. Nogot is a student in my class. He has told me that he owns a Ferrari, has shown me papers stating that he owns a Ferrari, and drives a Ferrari. Mr. Nogot owns a Ferrari.

Though the claimant wins the round understood as a round in the original justification game, he loses the round understood as a round in the ultra justification game because the ultracritic can disqualify the claimant's last move. Thus, the final move in the ultra justification game would be a move of the ultracritic indicating the false claim made by the claimant as follows:

Ultracritic: Eliminate your claim that Mr. Nogot owns a Ferrari! This is false!

The ultracritic wins the round.

Does the foregoing loss of the round to the ultracritic show that the Gettier problem has been solved simply by revealing that the claimant accepts something false? Is it merely false acceptances that defeat justification? The example in which the claimant's justification of this claim depends on his acceptance of the claim that Mr. Nogot owns a Ferrari may be dealt with in terms of the failure of the claimant to win the round in the game as illustrated above. That does not mean that the Gettier problem is solved, however.

Gettier noted that the justification of my claim that some student in my class owns a Ferrari need not be based on acceptance of the false claim that Mr. Nogot owns a Ferrari but might, instead, be based on the true claims that Mr. Nogot told me he owns a Ferrari, showed me papers stating he

owns a Ferrari, and drives a Ferrari.[3] In this case, the second move of the claimant in the justification game above would be as follows:

Claimant: It is more reasonable for me to accept that at least one student in my class who appears to own a Ferrari does own a Ferrari than to accept that none of the students in my class who appear to own Ferraris actually own Ferraris. (Mr. Nogot is a student in my class. He has told me that he owns a Ferrari, has shown me papers stating that he owns a Ferrari, and drives a Ferrari.)

Here the claimant wins the round in the ultra justification game, for there is nothing false in what he claims. This confirms Gettier's basic contention that justified true belief not based on any false acceptance may still fall short of knowledge. We shall be able to explain how the justification is defeated on our account, but to do so we must consider some feature of the ultrasystem and the ultra justification game beyond the mere elimination of false acceptances. The most obvious solution to the Gettier problem within the ultra justification game involving simply the denial of false acceptances does not succeed.

Some Inadequate Solutions

The most obvious solutions to the Gettier problem are in fact inadequate, for they depend on too simple an account of the role of truth and falsity in justification. It is nevertheless instructive to consider them, for they appeal to the most salient features of the original examples used to illustrate the problem. What we can learn from considering them is that an important problem often does not depend on the most salient features of the illustration of the problem.

For example, some philosophers aver that the problem is solved by requiring that for a person to know, her justification must not involve reasoning from a false premise.[4] Reasoning from a false premise, however, need not be involved, though it was in the original example. The reason is that a justification that a person has for accepting a true statement may be noninferential and not the result of reasoning, even though the justification is defeated by some false statement. An example from Chisholm, which we have already considered in another context, illustrates this quite clearly.[5] Suppose I see an object that looks exactly like a sheep and I, in fact, take it for a sheep. If I have considerable experience with sheep, I may be justified in accepting that I see a sheep. Imagine that I also see another object at the same time which does not look like a sheep, though it is one, and which I do not take for one. If the object I thus take to be a sheep is not one, then I do not know that I see a sheep, even though I am justified in accepting, and do accept, that I see one. Since the second object I see actually is a sheep, it is true that

I see a sheep. I have a justified true belief, but I do not know that I see a sheep because what I take for a sheep is not a sheep, and the sheep I see I do not take to be one. Here we do not have a reasoning at all. It is a simple case of mistaking one thing for another. It is an example of perceptual error.

Such examples lead some philosophers to demand that what justifies a person in accepting something must not justify her in accepting any false statement at all. But it may well be that whatever justifies us in accepting anything incidentally justifies us in accepting at least some false statements. Other philosophers have suggested that for a person to know something, she must not only be justified in accepting it but her justification must not contain any false statements or beliefs. This suggestion is, however, also inadequate.

A Harmless Error

Let us consider a small modification of the earlier example, in which Mr. Havit is replaced by a Mr. Knewit of whose Ferrari ownership I am knowledgeable. Suppose there are two men, Mr. Nogot and Mr. Knewit, each of whom I see before me with others in the room. Imagine that, from what I accept about Mr. Nogot, I am justified in accepting that he owns a Ferrari. Moreover, imagine that because of this, in response to the question of whether I know whether anyone in the room owns a Ferrari, I reply that I know that at least one person owns a Ferrari. Again, it seems that, if Mr. Nogot does not own a Ferrari but someone else in the room does, though I would have a justified true belief that at least one person in the room owns a Ferrari, I would not know this to be true. What justifies me in accepting this is my false belief that Mr. Nogot owns a Ferrari. Suppose, however, I am justified on independent and different grounds in accepting that Mr. Knewit owns a Ferrari and, indeed, Mr. Knewit does own a Ferrari. Suppose, for example, I sold Mr. Knewit my Ferrari! In this case, though part of what justifies me in accepting that at least one person in the room owns a Ferrari is my false belief that Mr. Nogot owns a Ferrari, I have a justification that does not depend on this false belief. It is based on my irrefutably justified and correct belief that Mr. Knewit owns a Ferrari.[6] Thus I am justified, although my justification contains a false statement or belief.

The Knowledge of Falsity

The foregoing illustrates why we have formulated condition *(iv)* in the analysis of knowledge as follows:

(iv) If *S* knows that *p*, then *S* is justified in accepting that *p* in some way that does not depend on any false statement

or, equivalently formulated in terms of defeasibility,

If *S* knows that *p*, then *S* is justified in accepting that *p* in some way that is not defeated by any false statement.

We require for irrefutable justification only that *S* has *some* justification that does not depend on any false statement or is not defeated by any false statement. We have yet to explain how a justification can be defeated by a false statement. A proposal advanced independently by Peter Klein and Risto Hilpinen is illuminating though defective.[7] In our terminology, they propose that the undefeated justification that a person *S* has for accepting that *p* depends on the false statement *q* if and only if *S* would not be justified in accepting that *p* if *S* knew *q* to be false. If I knew it to be false in the original Nogot and Havit case that Nogot owns a Ferrari, then I would not be justified in accepting that someone in my class owns a Ferrari. If I knew it to be false that what I take to be a sheep is a sheep, then I would not be justified in accepting that I see a sheep. On the other hand, in the last case considered concerning Nogot and Knewit, if I knew it to be false that Mr. Nogot owns a Ferrari, I would be justified in accepting that at least one person in my class owns a Ferrari because I am justified in accepting that Mr. Knewit owns one.

The Grabit Example

Nevertheless, the proposal is defective because it gives us an incorrect account of other examples. There are situations in which knowing some statement to be false would be misleading rather than clarifying. Another example from the literature illustrates this.[8] Suppose I see a man, Tom Grabit, with whom I am acquainted and have seen often before, standing a few yards from me in the library. I observe him take a book off the shelf and leave the library. I am justified in accepting that Tom Grabit took a book and, assuming he did take it, I know that he did.

Imagine, however, that Tom Grabit's father has, quite unknown to me, told someone that Tom was not in town today, but his identical twin brother, John, whom his father often confuses with Tom, is in town at the library getting a book. Had I known that Tom's father said this, I would not have been justified in accepting that I saw Tom Grabit take the book, for if Mr. Grabit confuses Tom for John, as he says, then I might surely have done so too. Under the Klein and Hilpinen proposal, I do not know that Tom Grabit took the book. The reason is this. Had I known it is false that Tom Grabit's father did not say what he did, I would have known that he did say what he did and, consequently, not have known that Tom Grabit took the book.

But do I lack knowledge? Notice that I originally have no reason to accept that Mr. Grabit, Tom's father, said what he did. Suppose, moreover, that Mr. Grabit made the remarks about John Grabit while raving alone in

his room in a mental hospital. The truth is that Tom's thieving ways have driven his father quite mad and caused him to form the delusion that Tom has a twin, John, who took the book from the library that was actually taken by Tom. Mr. Grabit thus protects his wish that Tom is honest. I know none of this, but I did see Tom Grabit take the book.

Mr. Grabit's remarks, of which I am totally ignorant, are completely misleading. We should, therefore, not deny that I know that Tom Grabit took the book because of the ravings of his father, of which I am fortunately ignorant. I know Tom Grabit took the book just as I know other people do the things I see them do. I accept nothing, moreover, concerning Tom's father and what he might or might not have said. Given these two features, I may be said to know that Tom Grabit took the book despite the fact that, had I known what his father said without knowing about his madness, I would not know whether it was Tom who took it.

The Newspaper Example

The Tom Grabit example is to be distinguished from one suggested by Gilbert Harman to illustrate how a person may lack knowledge, even though her belief is justified entirely by true statements.[9] Suppose a person, Ms. Readlucky, reads in a newspaper that a civil rights leader has been assassinated. The story is written by a dependable reporter who in fact witnessed and accurately reported the event. The reader of the story accepts this and is personally justified in accepting that the civil rights leader was assassinated. However, for the sake of avoiding a racial explosion, all other eyewitnesses to the event have agreed to deny that the assassination occurred and affirm that the civil rights leader is in good health. The newspaper then retracts the original story and reports that the civil rights leader is in good health as well. Imagine, finally, that all who are around Ms. Readlucky have, in addition to reading the story, heard the repeated denials of the assassination and thus do not know what to accept. But Ms. Readlucky knows nothing about these denials and has not read the retraction. Could we say that Ms. Readlucky, the one person who, by accident, has not heard the denials or read the retraction, knows that the civil rights leader was assassinated? The answer appears to be that she does not know.

If we agree to this, the obvious problem is to explain the difference between this case and the case of Tom Grabit. In both cases, there is some misleading information which, were it possessed, the person in question would not know. In the newspaper case, when a person lacks this information, we still deny that she knows that the civil rights leader was assassinated, whereas in the case of Tom Grabit, we affirm that the person knows that Tom Grabit took the book. What is the difference? In the Grabit example, no beliefs of mine concerning Tom's father or what he might have said serve to justify me in accepting that Tom took the book. In the newspaper

example, though this is unstated, part of what justifies the person in accepting that the civil rights leader has been assassinated is her belief that the newspaper is a trustworthy source of reliable eyewitness reports about the assassination. The person's justification for accepting that the civil rights leader has been assassinated depends on her false belief that the newspaper is a trustworthy source of reliable eyewitness reports about the assassination, but my belief that Tom Grabit took the book does not depend on any false belief. I do not have any belief about what Tom's father did or did not say. Moreover, my source of information about Tom—my senses and memory concerning what Tom is like—is trustworthy.

A Solution: Defeat and the Ultra Justification Game

The proper solution to these problems may be obtained from the ultra justification game by extending the role of the ultracritic in the game to include consideration of preferences and reasonings that supplement the acceptance system in the evaluation system of the claimant. Ultra justification, remember, requires justification without appeal to acceptances that are false, preferences for accepting something false over accepting something true, or reasonings that are unsound, for these are eliminated in what we have called the *ultrasystem* for the person because they are not compatible with a concern for truth. The ultrasystem, nevertheless, is required to acknowledge the existence of the eliminated states of acceptance, preference, and reasoning in the original evaluation system, for it is true that they are states of the person.

Let us see how this works in the examples we have thought about. Consider the original Gettier counterexample concerning Nogot and Havit and the resulting ultra justification game.

Claimant: Someone in my class owns a Ferrari.
Ultracritic: Nogot does not own a Ferrari.
Claimant: It is more reasonable for me to accept that Nogot owns a Ferrari than to accept that he does not own a Ferrari. I prefer accepting that Nogot owns a Ferrari to accepting that Nogot does not own a Ferrari because Nogot drives a Ferrari, says he owns it, and showed me papers of ownership, so Nogot owns a Ferrari.
Ultracritic: You must eliminate your claim that Nogot owns a Ferrari! (This is false!)

Notice that had the claimant appealed only to his preferences for accepting that Nogot owns a Ferrari to accepting that Nogot does not own a Ferrari without concluding that Nogot owns a Ferrari, the ultracritic could

again have won the justification game by demanding elimination of the preference of the claimant because his preference would be a preference for accepting something false over accepting something true. Thus the final comment of the ultracritic might have run as follows:

Ultracritic: You must eliminate your preference for accepting that Nogot owns a Ferrari to accepting that Nogot does not own a Ferrari! (It is false that Nogot owns a Ferrari!) You should not prefer accepting that Nogot owns a Ferrari over accepting that Nogot does not own a Ferrari. So it is not more reasonable for you to accept that Nogot owns a Ferrari than that he does not.

The claimant cannot win this round because the objection cited by the ultracritic cannot be answered or neutralized. The claimant has no information to support the claim that someone in her class owns a Ferrari. The claimant's justification is defeated.

Consider next the example in which the claimant is justified in accepting that Mr. Nogot owns a Ferrari, which is false, but is also justified in accepting that Mr. Knewit owns a Ferrari, which is true, and knows that Mr. Nogot and Mr. Knewit are students in her class.

Claimant: Someone in my class owns a Ferrari.
Ultracritic: Mr. Nogot does not own a Ferrari. No one in your class owns a Ferrari.
Claimant: It is more reasonable for me to accept that someone in my class owns a Ferrari than to accept that no one in my class owns a Ferrari. (Though Mr. Nogot does not own a Ferrari, Mr. Knewit does, as my information about him shows.)

This round is won by the claimant, and the claimant will remain victorious in the ultra justification game, showing the claimant to be justified in accepting that someone in her class owns a Ferrari in a way that is undefeated by the falsity of any statement the claimant accepts.

The example is like the original Nogot and Havit example except that the conclusion, someone in my class owns a Ferrari, instead of being inferred from the statement that Nogot owns a Ferrari, is inferred from the statements of evidence for accepting that Nogot owns a Ferrari. The reason is that the statements of evidence in question may all be true. Given our account of acceptance as a mental state having a certain functional role in inference, we may say that the claimant accepts the hypothetical to the effect that *if* the statements of evidence are true, then Nogot owns a Ferrari, even if, for some reason, the person does not accept the conclusion that Nogot owns a Ferrari.

The reason for ascribing acceptance of the hypothetical to the claimant is that the inference from the evidence to the conclusion that someone in the class owns a Ferrari rests on the acceptance of the hypothetical linking the evidence to that conclusion. Moreover, the acceptance of that hypothetical supports the preference for accepting that Nogot owns a Ferrari over accepting that he does not, which in turn supports the conclusion that it is more reasonable to accept that Nogot owns a Ferrari than that he does not. If it is not more reasonable to accept that Nogot owns a Ferrari than that he does not, then the objection that Nogot does not own a Ferrari is unanswered. Hence we can envisage the following ultra justification game:

Claimant: Someone in my class owns a Ferrari. If I have the evidence that Nogot owns a Ferrari, that he is student in my class, that he has told me that he owns a Ferrari, that he has shown me papers stating that he owns a Ferrari and that he drives a Ferrari, then Nogot owns a Ferrari. I have all that evidence consisting of true claims.

Ultracritic: You must eliminate your hypothetical claim that if you have the evidence that Nogot owns a Ferrari, that he is student in your class, that he has told you that he owns a Ferrari, that he has shown you papers stating that he owns a Ferrari and that he drives a Ferrari, then he owns a Ferrari! The evidence you have that Nogot owns a Ferrari is true, but Nogot does not own a Ferrari. It is not more reasonable for you to accept that someone in your class owns a Ferrari than that no one does. No one in your class owns a Ferrari.

The claimant loses the round because the objection cited by the ultracritic cannot be answered or neutralized. The justification is defeated.

The construction of the remaining ultra justification games to deal with other examples is left to the reader. In the case of Grabit, the claimant will win because there is nothing false that the person accepts which is relevant to the example, whereas in the case of the sheep the person falsely accepts that what she takes to be a sheep is a sheep and in the case of the assassination the person falsely accepts that the newspaper report is a trustworthy source of reliable eyewitness reports. These errors allow the ultracritic to win the ultra justification game, showing that the justifications are defeated.

One might worry that the ultracritic in the ultra justification games imagined above is merely a useless fiction because no one will actually be in a position to play the role of the ultracritic. This objection is, however, unwarranted. I may know enough about what another person accepts to play the winning role of an ultracritic against him. A person does not need to know the truth about everything a person accepts in order to know that some specific justification of hers depends on accepting some false claim,

for example, that Nogot owns a Ferrari. Since Nogot knows that Nogot does not own a Ferrari, he can play the winning role of the ultracritic against me, the claimant, in the original example concerning Nogot and Havit. Thus, we are often in a position to know that another person will lose the ultra justification game. In this way, the ultra justification game is a useful tool for evaluating the knowledge claims of others.

Truth Connection and the Isolation Objection

The appeal to the ultra justification game provides a reply to the most familiar objection to coherence theories. The objection is that coherence among a set of propositions or, in terms of the theory developed here, among members of the acceptance system, might fail to provide any connection with reality. The acceptance system and all that coheres with it could occur in a mind completely isolated from the external world. Internal coherence is not enough.

We may agree with the possibility of isolation but deny the actuality of it. Our reply to the objection is that undefeated justification reaches beyond internal coherence to external truth. Undefeated justification provides a truth connection between the mind and the world, between acceptance and reality. In fact, we may easily supply an argument that no form of the isolation objection can succeed against our theory. We may call it the *transformation argument*. Suppose that someone claims something about the world, that she sees a table in front of her, for example. To be personally justified, all objections must be answered or neutralized. Now, and this is the crucial point, the isolation objection is an objection. Consider the following justification game:

Claimant: I see a table in front of me.
Ultracritic: You are isolated from the external world.
Claimant: It is more reasonable for me to accept that I see a table in front of me than that I am isolated from the external world. (I am visually connected with the external world and not isolated from it.)

The evaluation system of the claimant enables her to answer the skeptical objection because she accepts that she is appropriately connected with the external world and not isolated from it. Assuming, finally, she is correct in accepting this, then her victory in this round of the justification game will be sustained in other rounds and transformed into a victory in the ultra justification game as well. Her justification will go undefeated.

If, on the other hand, she really is isolated from the external world and not visually connected as she accepts, perhaps because she is deceived by

the Cartesian demon, then she will lose the following ultra justification game:

Claimant: I see a table in front of me.
Ultracritic: You are not visually connected with the external world. You are isolated from the external world.
Claimant: I prefer accepting that I am visually connected to the external world and not isolated from it over accepting that I am not so connected and am isolated. Thus, it is more reasonable for me to accept that I am connected and not isolated than to accept that I am not connected and isolated. All my experience testifies to the existence of the external world.
Ultracritic: You must eliminate your claim that you see a table. You are not visually connected with the external world and are isolated. You must eliminate your preference for accepting that you are connected and not isolated over accepting that you are not connected and are isolated. It is not more reasonable for you to accept that you are connected and not isolated than that you are not connected and isolated.

The claimant loses the round because the objection cited by the ultracritic cannot be answered or neutralized. The justification is defeated.

Thus, the reply to the isolation objection is a dilemma. The claimant must accept that she is visually connected with external reality to win the justification game yielding personal justification. Either she is correct in accepting this, and she is so connected, or she is incorrect, and she is not connected. Suppose that she is connected with the external world as she accepts. In that case, she will be victorious in the ultra justification game, her justification will be undefeated, and she will turn out to have knowledge on our account. That is the appropriate result in such a case. Suppose, on the contrary, that she is not connected with the external world though she accepts that she is. Then she loses in the ultra justification game, and she will not turn out to have knowledge on our account. That is the proper result in such an instance, since she is truly ignorant. Whether she is isolated or not, our coherence theory of justification yields the appropriate result concerning whether she knows.

The conclusion is that victory in the ultra justification game ensures the appropriate truth connection between internal coherence and external reality. Such victory ensures this result because the objection that the connection is lacking is a skeptical objection. The objection must be met in terms of what one accepts and prefers to accept for one to be personally justified. So what one accepts must imply that one is appropriately connected. For personal justification to remain undefeated, it must be true that one is con-

nected in the way one accepts that one is. The truth connection transforms personal justification into knowledge.

Perception, Memory, and Introspection

The isolation objection is related to another kind of objection to the coherence theory of justification and knowledge. The objection is based on sources of knowledge favored by foundationalism as basic.[10] Perception, memory, and introspection are, it is alleged, basic sources of knowledge. It appears that we know that beliefs arising from these sources are justified because they originate in this way—because they are the products or outputs of our faculties—rather than because they cohere with some evaluation system or correction thereof. If I see something or remember something or introspect something, I appear to acquire a justified belief simply because it arises from sight, memory, or introspection regardless of what else I might happen to accept. Thus, the objection runs, we may have justified beliefs because they arise from some source or faculty of the mind independently of what else we accept and, therefore, coherence is inessential to the justified beliefs and knowledge emanating from them.

One sort of positive theory that generates an objection of this kind will be critically discussed in detail in the next chapter. It is, however, worthwhile to indicate how the objection can be answered in terms of the justification game and the ultra justification game that we have developed. Any claim that we see, remember, or introspect that something is the case immediately confronts the skeptical objection that the belief does not emanate from a trustworthy source for becoming informed of truth but arises in some untrustworthy manner. To be personally justified in accepting that one sees, remembers, or introspects something, one must, therefore, accept that these are trustworthy sources of information of the truth or the ultracritic will win the justification game.

Consider the following justification game concerning memory as an illustration:

Claimant: I remember drinking coffee for breakfast.
Ultracritic: Your memory is untrustworthy in the matter.
Claimant: It is more reasonable for me to accept that I remember drinking coffee for breakfast than that my memory is untrustworthy in the matter. My memory is clear and distinct, and memory of this sort is trustworthy.

The last sort of reply is required for the justification of the memory belief and depends on accepting that memory is trustworthy. It is not sufficient for knowledge that a belief arise from a trustworthy source when one has

no idea that this is so. The justification of the belief depends on acceptance of the unarticulated assumption that the source of the belief is trustworthy for becoming informed of truth and not deceptive. Victory in the justification game depends on our accepting that perception, memory, and introspection are trustworthy sources of information of truth in general and in our special circumstances. Victory in the ultra justification game depends in turn on our being correct in accepting these assumptions about the trustworthiness of our sources of information.

Personal justification of what we accept from perception, memory, and introspection results from our accepting that these sources are trustworthy to inform us of the truth, and personal justification is transformed into undefeated justification and knowledge only when we are correct in accepting that these sources are trustworthy to inform us of the truth. The transformation argument meets the objection, appealing to the sources of knowledge in the same way that it meets the isolation objection. Both objections give rise to a skeptical objection that must be met in order for a person to be personally justified. One objection is that one is isolated from the external world, and the other is that one's belief did not arise from a trustworthy source for informing us of the truth. What one accepts to meet such objections in order to obtain personal justification must turn out to be correct to yield undefeated justification. The correctness of what we accept about the sources and origins of the beliefs of perception, memory, and introspection transform such beliefs into knowledge. Coherence transforms sources of information into fountains of knowledge.

Knowledge of the Eliminated

Does the ultra justification game lead to the conclusion that we do not know that we accept something whenever it turns out to be false? To see the problem, return to the Nogot example. I do not know that Nogot owns a Ferrari, for the claim that he owns a Ferrari is false. But I do know that I *accept* that Nogot owns a Ferrari. When the falsity of my acceptance leads to the elimination of it in the ultra justification game, however, it appears that this might have the consequence that I do not know that I accept that Nogot owns a Ferrari.

However, this consequence does not result from the ultra justification game.[11] The reason is that the ultra justification game requires that the ultracritic acknowledge the existence of the states of acceptance, preference, and reasoning in the original justification game. Recall that the ultrasystem that is the basis of the ultracriticism used by the ultracritic contains not only the residual systems of acceptances, preferences, and reasonings once false acceptances, preferences for accepting falsehood over truth, and unsound reasonings are eliminated but also acknowledges the existence of the

states of acceptance, preference, and reasoning in the original evaluation system. What is true about the original evaluation system must be acknowledged by the ultracritic, for her role is only to expose error in the evaluation system. So, when the claimant affirms that he accepts something that he does accept and that is part of his original acceptance system, the ultracritic must acknowledge that he does accept it, and the objection that he does not accept it is answered by his acceptance in a way that is undefeated.

If the requirement that the states of the original system must be acknowledged in the ultrasystem and hence constrain the ultracritic seems a bias in favor of the claimant, it is important to notice that acknowledging the existence of such states also constrains the claimant and undermines the justification of some of his claims in the ultra justification game. Thus, suppose that in the example concerning Mr. Nogot and Mr. Knewit, where the claimant has the evidence that Mr. Knewit owns a Ferrari as a result of having himself sold Mr. Knewit a Ferrari, that the claimant is convinced that Mr. Knewit bought the car for someone else. In fact, Mr. Knewit bought the car for himself and owns it. So the claimant accepts that Mr. Knewit does not own the Ferrari the claimant sold to him, since he accepts that Mr. Knewit bought it for someone else. However, the claimant also accepts, for no good reason, that Mr. Knewit owns a Ferrari anyway. (The claimant accepts that no one would buy a Ferrari for anyone else unless he already owned one. Ferraris are just too wonderful!)

Clearly, the claimant does not know that someone in his class owns a Ferrari. He has evidence to accept that Mr. Nogot owns one, but, as the ultracritic will point out, that is false, and he has evidence to accept that Mr. Knewit owns a Ferrari, which is true. But the claimant accepts that Mr. Knewit does not own the Ferrari he sold to him because he thinks Mr. Knewit bought it for someone else. The claimant accepting that Mr. Knewit does not own a Ferrari will lead to his defeat in the ultra justification game. Thus, the ultracritic in the ultra justification game will win the game.

If the elimination of the false acceptance that Mr. Knewit does not own the Ferrari the claimant sold to him cleared the way for the claimant to justify the claim that Mr. Knewit owns a Ferrari on the basis of accepting that he sold Mr. Knewit a Ferrari, the claimant would have gained an illegitimate advantage in the ultra justification game for justifying the claim that someone in the class owns a Ferrari, which, of course, is true.

However, the original acceptance of the claimant to the effect that Mr. Knewit does not own the Ferrari he sold him must be acknowledged in the ultra justification game, and that suffices to ensure that if a claimant is not justified in accepting a target acceptance in the original justification, he will not be justified in accepting it in the ultra justification game either. Thus, the required acknowledgment of the existence of the states

of the original evaluation system, the acceptances, for example, is not a bias in favor of the claimant or his ultracritics but is simply the proper constraint on justification that converts to knowledge by remaining undefeated.

A Definition of Undefeated Justification

Having concluded that our conception of ultra justification yields an account of irrefutable and undefeated justification sufficient to solve the problems arising from the dependence of justification on error, let us give as precise a definition as possible of the notion of ultra justification. Consider the residual system, the truth compatible system, resulting from eliminating from the original evaluation system every false acceptance, every preference for accepting something false over something true and every unsound reasoning. Call this the t-system because it is the truth compatible subsystem of the original evaluation system. Now consider a system, which we shall call the *ultrasystem* of the person, containing the t-system with all the states of the system marked as t-states, t-acceptance, t-preferences, and t-reasonings combined with the other unmarked states of the original evaluation system. We restrict the use of ultrasystem in such a way that only the content of the t-states may be used to meet objections, answer them, or neutralize them, by the subject, the claimant, though the existence of all states of the original system must be acknowledged by both the claimant and the ultracritic.

In short, the existence of all states of the evaluation system must be acknowledged by both the claimant and the ultracritic, but only the content of those states that are t-states may be used to meet objections.

Thus, to illustrate with the original Nogot example, the acceptance of the claim that Mr. Nogot owns a Ferrari will be included in the ultrasystem but not marked as a t-state because it is false; therefore, the content of this unmarked acceptance, that Mr. Nogot owns a Ferrari, is excluded for use by the claimant in meeting objections of the ultracritic. By contrast, the acceptance of the claim that Nogot is a student in the class is marked as a t-state because it true, and the content of this acceptance, that Mr. Nogot is in the class, may be used by the claimant to meet objections of the ultracritic. Both the claimant and the ultracritic may use the claim that the claimant originally *accepted* that Mr. Nogot owns a Ferrari, as well as the claim that the claimant originally accepted that Mr. Nogot was a student in the class in the ultra justification game, because it is true that the claimant accepted both of these things and that truth must be acknowledged by both. It is only the content of the acceptance, that Mr. Nogot owns a Ferrari, that is excluded from use by the claimant in the ultra justification game.

We may then define undefeated justification as follows:

(*iv*d) *S* is justified in accepting that *p* in a way that is undefeated if and only if *S* is justified in accepting *p* on the basis of the ultrasystem of *S*.

Similarly, we may define what it means to say that the ultrasystem defeats a personal justification of *S* for accepting that *p* as follows:

The ultrasystem of S defeats the personal justification of *S* for accepting *p* if and only if *S* is personally justified in accepting *p* (justified on the evaluation system of S), but *S* is not justified in accepting *p* on the ultrasystem of *S*.

It should be noted that it is not necessary for a person to know what the members of her ultrasystem are in order to know that she has undefeated justification for accepting something and, therefore, to know that she knows. A person lifting her hand before her eyes accepts that she has a hand, and she also knows that her justification for accepting this does not depend on any error of hers. She might not know what the members of her ultrasystem are, but she does know that whatever they are, they will leave her justified in accepting that she has a hand. So, she may know and know that she knows. As a close analogy, a person may know that a set of theorems validly deduced from axioms contains no errors, even though she does not know exactly which theorems have been deduced, just because she knows the axioms are true and the person who deduced the theorems would not make any errors. A person can know that no correction of errors in her evaluation system will yield a system to defeat her justification for a specific claim because she knows that her justification for that claim is based on truths. It is not hard to know that you know when your evidence is good enough.

Knowledge Reduced to Undefeated Justification

The foregoing complicated set of definitions permits us to reduce knowledge to a simple formula. Knowledge is undefeated justified acceptance. The reduction is easily effected. We began in the first chapter with the following definition of knowledge:

DK. *S* knows that *p* if and only if (i) *S* accepts that *p*, (ii) it is true that *p*, (iii) *S* is justified in accepting that *p*, and (iv) *S* is justified in accepting

that *p* in a way that is not defeated by any false statement (that does not depend on any false statement).

We then undertook to analyze conditions (iii) and (iv) by means of a complicated set of definitions. The first definition specifies the system, the evaluation system, with which something must cohere to yield justification.

D1. A system *X* is an evaluation system of *S* if and only if *X* contains (a) states expressed by statements of the form, *S* accepts that *p*, attributing to *S* just those things that *S* accepts with the objective of accepting that *p* if and only if *p* (the acceptance system of *S*), (b) states expressed by statements of the form, *S* prefers accepting *p* to accepting *q*, attributing to *S* just those things that S prefers accepting with the same objective concerning acceptance, (the preference system of *S*), and (c) states expressed by statements of the form, *S* reasons from *p*, *q*, *r*, and so forth to conclusion *c*, attributing to *S* just those states of reasoning with the objective of being sound (having true premises and being valid).

The second definition expresses the idea that justification is coherence with a system.

D2. *S* is justified in accepting *p* at *t* on system *X* of *S* at *t* if and only if *p* coheres with *X* of *S* at *t*.

The next definitions articulate the idea that coherence with a system means that all skeptical objections can be met because they are either answered or neutralized on the basis of the system.

D3. *S* is justified in accepting *p* at *t* on system *X* of *S* at *t* if and only if all objections to *p* are answered or neutralized for *S* on *X* at *t*.

D4. *o* is an objection to *p* for *S* on *X* at *t* if and only if it is more reasonable for *S* to accept that *p* on the assumption that *o* is false than on the assumption that *o* is true, on *X* at *t*.

D5. An objection *o* to *p* is answered for *S* on *X* at *t* if and only if *o* is an objection to *p* for *S* on *X* at *t*, and it is more reasonable for *S* to accept *p* than to accept *o* on *X* at *t*.

D6. *n* neutralizes *o* as an objection to *p* for *S* on *X* at *t* if and only if *o* is an objection to *p* for *S* on *X* at *t*, the conjunction of *o* and *n* is not an objection to *p* for *S* on *X* at *t*, and it is as reasonable for *S* to accept the conjunction of *o* and *n* as to accept *o* alone on *X* at *t*.

We thus arrive at a definition of personal justification.

D7. *S* is personally justified in accepting that *p* at *t* if and only if *S* is justified in accepting that *p* on the basis of the evaluation system of *S* at *t*.

This is the required notion of justification used in (iii) in our definition of knowledge.

Finally, we require a definition of irrefutable or undefeated justification that amounts to a definition of the ultrasystem and justification on the basis of the ultrasystem for a person.

D8. A system *U* is the ultrasystem of *S* at *t* if and only if *U* contains as members (a) a truth-compatible subsystem of the evaluation system of *S* at *t*, a t-system, the states of which are marked as t-states, containing all states of *S* at *t*, which are states of acceptance wherein what is accepted is true, states of preference in which something false is not preferred to something true, and states of reasoning that are sound; and (b) the remaining unmarked states of the original evaluation system of *S* at *t*. The use of the system as a basis of justification is restricted so that only the content of the states of the t-system may be used to answer or neutralize objections, though the existence, but not the content, of unmarked states of the original evaluation system may also be used to formulate or meet objections.

D9. S is justified in accepting p in a way that is undefeated at *t (S* is irrefutably justified in accepting that *p* at *t)* if and only if *S* is justified in accepting *p* at *t* on the ultrasystem of *S* at *t*.

Needless to say, the attempt to analyze justification and undefeated justification in terms of acceptance, reasonableness, and truth has yielded a complicated analysis. As is often the case, however, thorough analysis enables us to find the underlying simplicity. We are now in a position to provide an elegant reduction of the original analysis of knowledge (DK). Knowledge reduces to undefeated justification, a just reward for our arduous analytical efforts.

The reduction of knowledge to undefeated justified acceptance is a consequence of our explication of condition (iv). This condition implies the other three. It is easiest to see that irrefutable or undefeated justification implies justification, that is, personal justification. If a person is justified in accepting that *p* on the basis of the ultrasystem, as our definition of irrefutable justification requires, the person is justified in accepting that *p* on the basis of the original evaluation system. The ultracritic can win any round that the critic can win in the original justification game. Thus, if a person's justification for accepting that *p* is irrefutable or undefeated, then the person is justified in accepting that *p*. Undefeated justified acceptance obviously implies acceptance, and the implication of the truth condition is trivial. If a person accepts

that p and it is false that p, then any justification the person has for accepting p will be refutable and defeated. The reason for this is that if it is false that p and the claimant accepts p, then the ultracritic in the ultra justification game may demand elimination of the acceptance that p. Thus, the claimant will lose the round in the game starting with the claim that p. This is equivalent to saying that if a person accepts that p when p is false, then the elimination of the acceptance from the ultrasystem will defeat the person's justification for accepting p. Hence, condition (iv), the defeasibility condition of (DK), our original definition of knowledge, logically implies the other three conditions, and knowledge is reduced to irrefutable or undefeated justification.

The reduction is a formal feature of the theory. The substance of it is the coherence theory of justification in which personal justification results from coherence with an evaluation system, just as the other necessary kind of justification, undefeated justification, results from modifications of the evaluation system. The soul of the theory is personal acceptance, preference, and reasoning. This is entirely an internal matter. One is personally justified in accepting something in part because what one accepts informs one that such acceptance is a trustworthy guide to truth. Even the conclusions that one accepts from perception and inference must cohere with one's background information articulated in an acceptance to ensure that they are trustworthy for being informed of the truth. Without such insurance, one may possess information but lack knowledge, for the trustworthiness of perception and inference is not a necessary a priori truth. When they prove trustworthy for being informed of the truth, this is the result of the nature of our faculties, the circumstances we find ourselves in and, most importantly, our background information about the circumstances in which our faculties are worthy of our trust. Personal justification requires acceptance of a body of *truth* to provide knowledge, however. A truth connection between acceptance and worldly fact is essential. There must be a match between what one accepts as a trustworthy guide to truth and what really is a trustworthy guide to truth. The match must be close enough to sustain justification when error is eliminated to obtain truth.

Given the importance of the trustworthiness of acceptance in yielding undefeated justification and knowledge, the theory might be regarded as a form of *reliabilism* (which we consider in the next chapter),[12] but given that the acceptance of our trustworthiness yields, in the normal case, justification of its own acceptance, the theory might as well be called foundational coherentism.[13] To obtain knowledge we need the right mix of internal and external factors. Our theory may appear dialectically promiscuous, but fidelity to a single approach is puritanic oversimplification. The simple theory, though

ever seductive, is usually the mistress of error. The queen of truth is a more complicated woman but of better philosophical parts.

Determining Justification

How can a person ascertain that she is irrefutably justified in accepting that *p* or that her justification is undefeated except by checking to determine whether she is personally justified in accepting that *p*? At a given point in time, a person can only evaluate such a claim in terms of her acceptance system at that time. Her acceptance system summarizes her information; it is her repository of information about the world, and it alone is the basis for evaluation. That is the fundamental truth in the subjective approach. The acceptance system is the first and last court of appeal of an individual at any specific point in time. It is the only epistemic court for synchronic adjudication—judgment at a given time.

We are, however, diachronic creatures spanning time and considering our cognitive accomplishments through time. As a result, an individual remembers that, however many successes he has had, he also has had his failures. In the past, some of the things that he accepted were false and, therefore, at those times there was a distinction between his acceptance system and his ultrasystem. He may take what steps he can to make the two coincide as closely as possible in the future, of course, but when the time for evaluation of various claims to truth arrives, the basis for evaluation must be the system he has then. When I consider whether I am irrefutably justified in accepting something now, I can only determine whether I am personally justified and, if I am, conclude that I am irrefutably justified. Others with the information about where I have erred may disagree, and I may come to agree with them at some later date when further information has won my acceptance.

We have thus arrived at a coherence theory of justification. Coherence with an evaluation system yields personal justification, and the addition of coherence with an ultrasystem yields undefeated justification. Coherence with the ultrasystem keeps the justification undefeated. Knowledge, or undefeated justification, results from the right combination of coherence, acceptance, preference concerning acceptance, reasoning, and truth. We may put the matter this way: We accept what we do with the objective of reaching truth and avoiding error. What we are personally justified in accepting depends on what we accept, what we prefer to accept, and how we reason with these objectives. Whether our justification is undefeated depends on whether we succeed in our attempt. If we win the original justification game without depending on error, our justification is irrefutable or undefeated and we gain knowledge. If our victory rests on error, we have won a game of justification but lost the prize of knowledge.

Introduction to the Literature

The original very readable brief article that spawned a vast literature is Edmund Gettier's "Is Justified True Belief Knowledge?" in which Gettier argued for the negative. There are many important articles written to deal with the problem that Gettier raised. One excellent collection is *Knowing: Essays in the Analysis of Knowledge,* edited by Michael Roth and Leon Galis. Another important collection is edited by George S. Pappas and Marshall Swain, *Essays on Knowledge and Justification.* For those who become fascinated with the problem, the literature is summarized ultracritically in *The Analysis of Knowing,* by Robert K. Shope.

Notes

1. See Donald Davidson, "Truth and Meaning" and "Radical Interpretation," in his *Inquiries into Truth and Interpretation* (Oxford: Clarendon, 1984), esp. 27, 136–37. See also Davidson's "A Coherence Theory of Truth and Knowledge," in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson,* ed. Ernest LePore (Oxford: Blackwell, 1986), 307–19. Davidson's claims have been cogently criticized by Peter Klein, "Radical Interpretation and Global Skepticism," *Truth and Interpretation,* 369–86; and Ernest Sosa, "'Circular' Coherence and 'Absurd' Foundations," in *Truth and Interpretation,* 387–97.

2. Edmund Gettier Jr., "Is Justified True Belief Knowledge?" *Analysis* 23 (1963): 121–23. The example in the text is taken from Keith Lehrer, "Knowledge, Truth and Evidence," *Analysis* 25 (1965): 168–75.

3. Gettier made an observation to this effect in a symposium of the Eastern Division meetings of the American Philosophical Association, December 1970, in reply to Gilbert Harman's "Knowledge, Reasons, and Causes," *Journal of Philosophy* 67 (1970): 841–55.

4. See Irving Thalberg, "In Defense of Justified True Belief," *Journal of Philosophy* 66 (1969): 794–803; and Joseph Margolis, "The Problem of Justified Belief," *Philosophical Studies* 23 (1972): 405–09.

5. R. M. Chisholm, *Theory of Knowledge* (Englewood Cliffs, N.J.: Prentice-Hall, 1966), 23 n. 22.

6. This argument appeared originally in the author's "Knowledge, Truth, and Evidence," *Analysis* 25 (1965): 168–75, and was reprinted in Michael Roth and Leon Galis, eds., *Knowing: Essays in the Theory of Knowledge* (New York: Random House, 1970), 55–66.

7. Peter Klein, "A Proposed Definition of Propositional Knowledge," *Journal of Philosophy* 67 (1971): 471–82; and Risto Hilpinen, "Knowledge and Justification," *Ajatus* 33 (1971): 7–39.

8. Keith Lehrer and Thomas Paxson Jr., "Knowledge: Undefeated Justified True Belief," *Journal of Philosophy* 66 (1966): 225–37.

9. Gilbert Harman, *Thought* (Princeton: Princeton University Press, 1973).

10. Robert Audi, *Belief, Justification, and Knowledge* (Belmont, Calif.: Wadsworth, 1988).

11. I am indebted to Jonathan Tweedale for the discussion that follows and to his paper, "Epistemic Humility and Undefeated Justification," forthcoming in *STOA* (2000).

12. Cf. Alvin Goldman, *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge: MIT Press, 1992).

13. Cf. Susan Haack, *Evidence and Inquiry: Towards Reconstruction in Epistemology* (Oxford: Blackwell, 1993).

This page intentionally left blank

8

Externalism and the Truth Connection

Our analysis of irrefutable and undefeated justification in terms of coherence and truth within an acceptance system brings us into conflict with an important competing theory of knowledge called *externalism.* The fundamental doctrine of externalism is that what must be added to true belief to obtain knowledge is the appropriate connection between belief and truth. An earlier account presented by Alvin Goldman affirmed that the appropriate connection is causal.[1]

This is a very plausible sort of account of perceptual knowledge. The fact that I see something, the hand I hold before me, for example, causes me to believe that I see a hand. The fact that my seeing a hand causes me to believe I see a hand results, it is claimed, in my knowing that I see a hand. According to such an analysis, it is the history of my belief, a matter of external causation, rather than coherence with some internal system, that yields knowledge. The central tenet of externalism is that some relationship to the external world accounting for the truth of our belief suffices to convert true belief to knowledge without our having any idea of that relationship. It is not our conception of how we are related to a fact that yields knowledge but simply our being so related to it.

The early analysis, though providing a plausible account of perceptual knowledge, was a less plausible account of our knowledge of generalities—that men do not become pregnant, for example, or that the atom has been split or that there is no largest prime number. For here the nature of the required causal relationship between what is believed and the belief of it

evades explication. That objection to externalism is, however, one concerning only a specific version thereof and leaves open the possibility that other versions might succeed.

Later analyses by others, and by Goldman himself, aim at preserving the thesis of externalism that some relationship of the belief to what makes it true yields knowledge, whether we have any idea of that relationship or not.[2] D. M. Armstrong and Fred Dretske have argued that the relationship should be construed as nomological, one resulting from some law of nature connecting the belief with what makes it true.[3] This account is closely connected with Robert Nozick's proposal that belief track truth is in a sense explicated, in part, by the counterfactual claim that the person would not have believed what she did if it were not for the truth of the belief.[4] Goldman now claims that justified belief, which he takes as a necessary condition of knowledge, must be the result of a belief-forming process that reliably yields truth.[5] Other externalists deny that justification is necessary for knowledge. They all agree, however, that a belief resulting from a certain kind of process or relationship connecting beliefs with truth can convert them to knowledge without the sustenance or support of any other beliefs or system of beliefs.

Naturalism

Assuming that the required relationship is something like causation, such externalist theories are claimed to be *naturalistic.* What is a naturalistic theory? It is one in which all the terms used in the analysis are ones that describe phenomena of nature, such as causation, for example, or that can be reduced to such terms. Hume's theory of belief was naturalistic in this sense. He restricted his account of human knowledge to relations of causation, contiguity, and resemblance.[6] It was W. V. O. Quine, however, who introduced the term *epistemology naturalized* and suggested that inquiry into the nature of human knowledge be restricted to accounts of how belief arises and is altered.[7] Other philosophers have adopted the term to refer simply to all accounts of knowledge that are couched in naturalistic vocabulary or reducible to such a vocabulary. The early account by Goldman considered above, according to which *S* knows that *p* if and only if *S*'s believing that *p* is caused in the appropriate way by the fact that *p,* is in this extended sense an example of epistemology naturalized. Other early naturalistic accounts offered by Armstrong and Dretske rested on the assumption that the conversion relation was based on nomological rather than causal relations, that is, relations articulated in laws of nature.[8] Dretske's basic idea was that the reasons we have for believing what we believe should be nomologically connected with the truth of what is believed, that is, that it should be a law of nature that a person having such reasons for

believing what she does will have a true belief. Assuming a naturalistic account of having a reason, which Dretske supplies, such an account is also naturalistic.

One interesting aspect of some externalistic theories that naturalize epistemology is the way in which they attempt to avoid the problems of foundationalism. According to Dretske or Nozick, for example, there is no need either to justify beliefs or posit self-justified beliefs because, contrary to the traditional analysis, the justification of beliefs is not required to convert true beliefs into knowledge. Beliefs or true beliefs having the appropriate sort of naturalistic external relationships to the facts are, as a result of such relationships, converted into knowledge without being justified. It is the way true beliefs are connected to the world that makes them knowledge rather than the way in which we might attempt to justify them. Notice how plausible this seems for perceptual beliefs. My belief that I see a bird is caused by my seeing a bird, and that accounts for my knowing that I see a bird rather than some justification I have for that belief. What matters for knowledge is how the belief arises, not how I might reason on behalf of it. The traditional analysis says that knowledge is true belief coupled with the right sort of justification. One sort of externalist analysis says that knowledge is true belief coupled with the right sort of naturalistic relation. It is plausible to assume that the naturalistic relationship will be one concerning how the belief arises, in short, the natural history of the belief. Looked at in this way, the justification requirement can be eliminated altogether in favor of the right sort of historical account.

The Advantages of Externalism

Before turning to details and objections, we shall note the advantages of externalism. First, according to some externalists, the need for justification and a theory of justification is eliminated as a component of an analysis of knowledge. On such an account, it is admitted that inference may play some role in the natural history and reliable formation of a true belief, but it is also possible to hold that some beliefs are noninferential and yet constitute knowledge. They are beliefs arising from experience in the naturalistically appropriate way without the intervention of inference. This may be offered as an account of what the foundationalist was searching for, but in the wrong place. True beliefs that arise in the appropriate way from experience are knowledge because of the way they arise. There is no need to affirm that such beliefs are self-justified to maintain that they convert to knowledge. We might think of such beliefs as naturalized basic beliefs. Such basic beliefs might, of course, serve as the premises for inferring other beliefs, and such inference might convert those beliefs to knowledge as well. It

is the history of the belief rather than some sort of justification of the belief that converts it to knowledge.

A Reply to Skepticism

It is helpful, as well, to notice how neatly this sort of theory deals with traditional and modern forms of skepticism. The skeptic, confronted with a commonsense perceptual claim, that I see a tree, for example, has traditionally raised some skeptical doubt, the Cartesian one, for example, that we might be deceived by an evil demon who supplies us with deceptive sensations that lead us to believe we see external objects when we do not see them at all. Or consider the case of a small object, a 'braino,' implanted in our brain which, when operated by a computer, provides us with sensory states that are all produced by the computer influencing the brain rather than by the external objects we believe to exist.[9] In neither case, affirms the skeptic, do I know I see a tree when I believe I do. I only know that I see a tree if I know that it is not the demon or the braino that produces my belief, the skeptic avers, and insists that I do not know this. Why do I not know that there is no demon or braino? I do not know so because my experience would be exactly the same if there were; that is what the demon and braino do, produce experiences that exactly duplicate the experiences I would have if I were to see a tree. I have no evidence whatever from experience against these skeptical hypotheses and, therefore, the skeptic concludes, I do not know them to be false.

The externalist reply is simple. If my beliefs are, indeed, produced by the demon or by the braino, then they are false and I am ignorant. On the other hand, if the beliefs are true and produced in the appropriate way, then I do know. I do not need to *know* that the skeptical hypotheses are false to know that I see a tree, though, of course, the skeptical hypotheses must *be* false. Otherwise, my belief that I see a tree will be false. All that is necessary is that my belief be true and that it arise in the appropriate way—that it have a suitable history—for it to constitute knowledge. If my belief is true and has arisen in the appropriate way, then I know that I see a tree, even if I do not know that the conflicting skeptical hypotheses are false. I might never have considered such skeptical machinations. Confronted with them, I might be astounded by them and find them so bizarre that they are not worthy of consideration.

The skeptic might retort that I cannot so easily escape the clutches of skepticism. For example, she might suggest that when I claim to know that I am seeing a car, a Mazda RX7, for example, I must have the information required to tell a Mazda RX7 from cars of another sort and, lacking such information, I do not know that I see a Mazda RX7. Hence, I must know that the car is not a Toyota MR2 or a Porsche 944, which bear some resemblance to a Mazda RX7. Going on, the skeptic might argue that to know

that I see a Mazda RX7, I must have the information required to distinguish seeing a Mazda RX7 from experiences of another sort, those supplied by the demon or braino, and, lacking such information, I do not know that I am seeing a Mazda RX7 or even that I am seeing a car. So, the skeptic concludes, just as I must know that the car I am seeing is not of another manufacture, so I must know that my experiences are not of skeptical manufacture. That, she insists, is precisely what I do not know. Skepticism wins.

Relevant Alternatives: A Reply to the Skeptic

The reply of the externalist is a combination of counterassertion and explanation. The counterassertion is that my true belief that I see a tree arising in the way it does is knowledge, even if I do not know that it has arisen in that way rather than in the way the skeptic suggests. If the skeptical hypothesis is true and the belief has not arisen in the way I suppose, then I lack knowledge; but if it has arisen in the way I suppose, then I have knowledge, even if I do not know competing hypotheses about the origin of the belief to be false. It does not matter whether I know that the belief originated in the appropriate manner. All that matters is that it did originate in that way. Then I know.

Moreover, the reason it is unnecessary to rule out our competing hypotheses is that they are irrelevant. The explanation about the Mazda, for example, is that there will be some cases, but not all, in which some information excluding other alternatives will be necessary for knowledge. The alternative that I am seeing a Porsche 944 and not a Mazda RX7 is a relevant alternative. The alternative that I am being deceived by an evil demon or a braino is not.[10] What is the difference? My information about what a Mazda RX7 looks like must be sufficient to enable me to distinguish it from other cars, and that information plays a role in the formation of my belief that I am seeing a Mazda RX7. In other cases, particularly those suggested by the skeptic in which there is no such distinguishing information, no such information enters into the appropriate origination of the belief. Where the distinguishing information is a necessary component in the suitable generation of the belief, the alternatives to be distinguished from the truth are relevant, but where it is not a necessary component, the alternatives are not relevant ones.

A skeptic might find the distinction between relevant and irrelevant alternatives capricious and question-begging as a counterargument. Nevertheless, the initial reply to the skeptic to the effect that true belief originating in the appropriate manner is knowledge, even if we do not know the skeptical hypotheses to be false, is a straightforward consequence of epistemology naturalized, whether or not it satisfies the demands of the skeptic.

The Naturalistic Relation

The advantages of naturalism are robust, but the theory must be true, not merely advantageous, to solve the problems with which we began. To ascertain whether the theory is true, we must have some account of the naturalistic relationship that is supposed to convert true belief into knowledge. One might be a naturalist about justification and maintain that justification is reducible to some naturalistic relationship. In fact, a philosopher eager to connect the naturalistic analysis with the traditional analysis might argue that a person has the requisite sort of justification for knowledge if and only if true belief arises in the appropriate naturalistic manner. This would provide us with a naturalistic reduction of justification. Thus, the externalist theory can be construed as a naturalistic account of justification or as a repudiation of a nonnaturalistic account of justification. As we shall see later, however, there are objections to externalist accounts of justification that might lead an externalist to prefer the repudiation strategy.

What exactly is the external relationship that converts true belief into knowledge? It is typical of epistemological theories to take some sort of example as a paradigm of knowledge, to fine-tune the theory to fit that sort of example and, at least at the outset, to ignore less felicitous examples whose subsequent consideration necessitates rather substantial modification of the theory. That is the history of externalism.

The paradigm example for the externalist is perception. In the case of perception, it is indeed very plausible to contend that what converts perceptual belief into knowledge is the way that the belief arises in perceptual experience. My belief that I see a tree is converted into knowledge by being caused by my actually seeing a tree. Another kind of example is communication. You tell me that Holly Smith is Dean and that causes me to believe that Holly Smith is Dean. Do I know that Holly Smith is Dean as a result of this causation? It might be contended, and has been, that if my informant knows that what he tells me is true, then I know because he knows and his communication caused me to believe this. Of course, his knowing remains to be explicated. The assumption is that there is a causal chain beginning with the fact that Holly Smith is Dean and ending with my believing it, which accounts for my knowing it.

Thus, following Goldman's early proposal, we might consider the following as characteristic of externalistic theories that eliminate the justification condition.

(CK) *S* knows that *p* if and only if *S* believes that *p* and this belief is caused in the appropriate way by the fact that *p*.[11]

This account leaves us with the need to explain the difference between being caused in an appropriate way and being caused in a way that is not appropriate. Typical cases of perception provide a model of the appropriate kind of causation.

Dretske has suggested that when x is something S perceives, then

(DK) S knows that x is F if and only if S's belief that x is F is caused or causally sustained by the information that x is F received from the source x by S.[12]

Dretske's analysis, though restricted to perceptual knowledge, highlights two needed qualifications recognized by other authors as well. The first is that the belief need not be caused but only causally sustained by the information that p. This is necessary because the originating causation of a belief might involve an error that is corrected by subsequent information one receives.

For example, if I see two men in the distance, I might take the one on the left to be Buchanan and believe that I see Buchanan when, in fact, it is not Buchanan, as I note when I move closer, but Harnish instead. At the same time, I note that the other man, the one on the right, is Buchanan and that Buchanan and Harnish are dressed in such a way that each appears to be the other in preparation for Tolliver's Halloween party. My belief that I see Buchanan was caused by my seeing Harnish dressed as Buchanan, and I continue to hold that belief subsequently when I receive the further information that corrects my mistake about the man on the right but sustains my belief that I see Buchanan and, indeed, that I saw him earlier, though I did not recognize him. Moreover, on this sort of account the appropriate kind of causal relation is explicated in terms of receiving information from a source.

The foregoing analyses are, however, too restricted in scope to provide us with a general analysis of knowledge. There is more to knowledge than perceptual knowledge, and not all knowledge that p can be supposed to be caused by the fact that p. The most obvious example is general knowledge, my knowledge that all human beings die, for example. That fact includes the fact of death of as yet unborn humans, which cannot now cause me to believe that all humans die or causally sustain that belief. Our knowledge that all neutrinos have minute rest mass is yet more difficult to account for on such a model, since no one has ever perceived a neutrino at rest. Assuming there to be mathematical knowledge, for example, that integers are infinite, the causal theory seems inappropriate. The integers appear to lie outside the temporal order and to be incapable of causing anything.

Accounts of knowledge in terms of causation or the receipt of information fail to provide an account of our knowledge of general and theoretical truths. Moreover, it is easy to see that externalism in no way requires such a restrictive conception of the external relationship. Causal, or information-receiving, analyses of knowledge have the virtue of explicating knowledge in a way that explains the connection between truth and belief—between reality and thought—and provides an answer to skepticism. We may, however, maintain the connection between truth and belief without committing ourselves to a restrictive causal connection. Instead, we may require that the *history* of the belief connect the belief with truth.

There are three popular accounts of how the history of a belief might connect the belief with truth. The first and perhaps best known is the later account of Goldman, according to which true belief is converted to knowledge via justification when the belief is the result of a reliable belief-forming process. Goldman's basic idea, which he has modified and refined, is as follows:

> If *S*'s believing that *p* at *t* results from a reliable cognitive belief-forming process (or set of processes), then *S*'s belief in *p* at *t* is justified.

The refinements include an account of reliable rules, methods, and processes.[13] The second account, offered by Nozick, who rejects a justification condition, requires that a belief track truth in order to convert to knowledge in the sense that the person would believe that *p* if *p* were true and would not believe that *p* if *p* were not true.[14]

The third account is due to Plantinga, who also rejects a justification condition. His analysis of knowledge includes a number of components, but his most basic idea is that what converts a belief to knowledge is the belief's resulting from a properly functioning cognitive faculty that is successfully designed for the purpose of producing true beliefs in the present environment.[15] The notion of a properly functioning faculty, like that of a reliable belief-producing process, is historical. What is necessary and, given the appropriate other conditions, sufficient for conversion to knowledge is how the belief was produced for the purpose of yielding true beliefs. Plantinga argues for the importance of design for this purpose and concludes that theism and not evolution is required to explain such design. His premise is that God and not evolution can be expected to design us in a way that our faculties would produce true beliefs. A theist might think that the goodness of God would ensure such a design, but a detractor might insist that the obvious presence of evil in the world might include the evil of our cognitive faculties' being deceptive. Whatever the outcome of that dispute, our concern in the present chapter is the claim Plantinga shares with the naturalist

philosophers that a belief converts to knowledge because of how it is produced—on his account, by a properly functioning cognitive faculty.

The three theories share some advantages. They retain the basic reply to the skeptic considered above, namely, that an external relationship suffices for knowledge. They accomplish this without assuming that we have any guarantee that our beliefs are true, moreover. That my belief is the outcome of a reliable belief-forming process does not presuppose that I have any guarantee of the truth of the belief. Similarly, I might believe that something is true when I would not have believed it had it not been true, even though I have no guarantee that it is true. Finally, my belief might be produced by a properly functioning cognitive faculty aimed at producing true beliefs, although I have no guarantee that my belief is true. Thus, given either account of knowledge, the skeptic may be answered while allowing, what seems obvious, that we are fallible in the way in which we form our beliefs, even those converting to knowledge. The result is a fallibilistic epistemology without the postulation of self-justified beliefs.

Objections to Externalism: Information Without Knowledge and the Opacity Objection

There is, however, a general objection to all externalist theories that is as simple to state as it is fundamental: the external relationship might be opaque to the subject, who has no idea that her beliefs are produced, caused, or causally sustained by a reliable belief-forming process or properly functioning cognitive faculty. The person might fail to know because of the opacity to her of the external relationship and her ignorance of it. Any externalist account faces the fundamental opacity objection that a person totally ignorant of the external factors connecting her belief with truth might be ignorant of the truth of her belief as a result. All externalist theories share a common defect, to wit, that they provide accounts of the possession of information, which may be opaque to the subject, rather than of the attainment of transparent knowledge. The appeal of such theories is their externalistic character. They assimilate knowledge to other external relationships between objects. Our attainment of knowledge is just one external natural relationship between facts among all the rest. It is a relationship of causality, or nomological correlation or frequency correlation or counterfactual dependence. This very feature of such theories, which makes them so appealing, is their downfall. The relationship in question may suffice for the recording of information, but if we are ignorant of the relationship, if it is opaque to us, then we lack knowledge.

As in our refutation of foundationalism, what is missing from the accounts of externalists is the needed supplementation of background infor-

mation and the transparency of it. To ensure that the specified relationships convert into knowledge, we need the additional information. Information about the existence of those relationships making them cognitively transparent to the subject might suffice for the conversion into knowledge, as might information about crucial effects of the relationships. For example, information about the truth frequency of a kind of belief resulting from a reliable process might suffice even if the person lacked information about the process itself. Additional information about how either the external relationship or the effect of high truth frequency is necessary, however, for the conversion of the belief into knowledge. Such additional information is what is needed to supplement the information contained in the belief alone, and it is precisely the sort of information required for coherence and irrefutable justification.

The general opacity problem for externalism can be seen most graphically by considering an analogy proposed by Armstrong. He suggested that the right model of knowledge is a thermometer.[16] The relationship between the reading on a thermometer and the temperature of the object illustrates the theories mentioned above. Suppose that the thermometer is an accurate one and that it records a temperature of 104 degrees for some oil it is used to measure. We can say, with Armstrong, that there is a nomological connection between the temperature and the thermometer reading; with Dretske that the thermometer receives the information; with Nozick that the thermometer would not record a temperature of 104 degrees if it were not true that the oil was at 104 degrees; and with Goldman that the reading is the outcome of a reliable temperature-recording process. The problem with the analogy is that the thermometer is obviously ignorant of the temperature it records. The question is—why?

One might be inclined to suggest that the thermometer is ignorant of temperature only because it lacks the capacity of thought. If, contrary to fact, the thermometer could entertain the thought that the oil is 104 degrees, would that suffice? Would the thermometer know that the temperature is 104 degrees? What are we to say of this fanciful thought experiment? One might protest, of course, that it is too far-fetched to turn the philosophical lathe. The thermometer does record information accurately, however, and, given the imagined capacity for thought, it may be said that the thermometer possesses that information as well. But our thoughtful thermometer does not *know* that the temperature of the oil is 104 degrees as a result of having the thought that this is so. The reason is that it might have no idea that its thoughts are accurate temperature-recording thoughts. If it has no idea that this is so, then, even if it has the thought that the temperature of the oil is 104 degrees when it records that temperature, it has no idea whether the thought is correct. The correctness of the thought is opaque to it. To obtain the benefits of these reflections, however, it is necessary to move to the hu-

man case to fully appreciate the implications of opacity.

Suppose a person, Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device that is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, is implanted in Truetemp's head so that the very tip of the device, no larger than the head of a pin, sits unnoticed on his scalp and acts as a sensor to transmit information about the temperature to the computational system in his brain. This device, in turn, sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process and a properly functioning cognitive faculty.

Now imagine, finally, that Mr. Truetemp has no idea that the tempucomp has been inserted in his brain and is only slightly puzzled about why he thinks so obsessively about the temperature; but he never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. Thus, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is? Surely not. He has no idea whether he or his thoughts about the temperature are reliable. What he accepts, that the temperature is 104 degrees, is correct, but he does not know that his thought is correct. His thought that the temperature is 104 degrees is correct information, but he does not know this. Though he records the information because of the operations of the tempucomp, he is ignorant of the facts about the tempucomp and about his temperature-telling reliability. Yet the sort of causal, nomological, statistical, or counterfactual relationships required by externalism may all be present. Does he know that the temperature is 104 degrees when the thought occurs to him while strolling in Pima Canyon? He has no idea why the thought occurred to him or that such thoughts are almost always correct. He does not, consequently, know that the temperature is 104 degrees when that thought occurs to him. The correctness of the thought is opaque to him.

It might be useful to add a bit to the story to reinforce the conclusion that Mr. Truetemp does not know that the temperature is 104 degrees when the tempucomp causes him to have the thought that the temperature is 104 degrees which he accepts. The addition will, at the same time, make it clear that the tempucomp is designed with the benevolent purpose of producing true beliefs when it is functioning properly. Let us add that the doctor who installed the device during brain surgery thought it useful to install the device because of a malady Mr. Truetemp has that can trigger a fatal stroke if he is outside when the temperature hits 104 degrees. Moreover, Mr. Truetemp is aware of this danger. However, the tempucomp is only an ex-

perimental device that was untested when implanted. So Mr. Truetemp's surgeon, though he hoped the device might work, did not know whether it would produce the effect for which it was designed. The point is that not only Mr. Truetemp is ignorant of the reliability of his thoughts that the temperature is 104 degrees, but so is everyone else. No one knows that he is having true thoughts when he has thoughts that the temperature is 104 degrees. The correctness of these thoughts is opaque to all.

The preceding example is not presented as a decisive objection against externalism and should not be taken as such. Both Goldman and Plantinga have replies to the objection appealing to further conditions on what converts belief to knowledge. One can place some constraint on the relationships, processes, or faculties converting belief to knowledge to exclude production by the tempucomp as a case of knowledge. The fundamental difficulty remains, however—more than the possession of correct information is required for knowledge. One must have some way of knowing that the information is correct. Consider another example. Someone informs me that Professor Haller is in my office. Suppose I have no idea whether the person telling me this is trustworthy. Even if the information I receive is correct and I believe what I am told, I do not know that Haller is in my office because I have no idea of whether the source of my information is trustworthy. The trustworthiness of the information and of the source of the information is opaque to me. Opacity deprives me of knowledge. The nomological, statistical, or counterfactual relationships or processes may be trustworthy, but this is not transparent to me, for I lack information about them.

When we considered the distinction between belief and acceptance in the third chapter, we noted the argument to the effect that a person who receives the information that p and believes that p as a result may fail to know that p. The reason is that the person may not know that the information she thus receives and believes is correct information. If a person does not know that the information, that p, which she receives is correct information, then she does not know that p. All forms of externalism fail to deal with this problem adequately. To know that the information one possesses is correct, one requires background information about that information. One requires information about whether the received information is correct or not, and lacking such information, one falls short of knowledge. The correctness of the information is opaque to one. This is a line of argumentation we have already encountered in earlier chapters. A necessary condition of knowledge is coherence with background information, with an acceptance system, informing us of the correctness of the information we possess and rendering the correctness transparent to us.

Externalism and Justification

Some externalist theories repudiate justification as a condition of knowledge, those of Nozick and Dretske, for example.[17] Such theories may provide an interesting account of what it is like for belief to constitute correct information or to track truth, but they provide no account of knowledge. The reason is that no one knows that what she accepts is true when it would have been just as reasonable for her to have accepted the opposite on the basis of her information. A necessary normative condition of a person knowing that p is that it be more reasonable for her to accept that p than to accept the denial of p on the basis of her information. This condition implies the need for a justification condition of the sort we have proposed.

One may, as Goldman illustrates, combine externalism with the affirmation of a justification condition, but such an account, if it considers background information in an acceptable manner, will introduce a coherence factor. Goldman insists, for example, that a justified belief resulting from a reliable belief-forming process must not be undermined by other evidence the subject possesses.[18] The condition requiring that the belief not be undermined by other evidence is a kind of negative coherence condition to the effect that the belief not be incoherent with background information. Nevertheless, the source of justification on this account is the reliability of the belief-forming process, that is, the fact that the belief has the sort of history frequently producing true beliefs. As a result of providing a justification condition, a normative constraint is supplied by Goldman's account.

The objection raised against externalism in general still applies to such a theory, however. A person totally ignorant of the reliability of the process producing his belief might not know that what he believes is true because of lack of information about the correctness of the belief, for example, about the truth frequency of beliefs of the kind in question, even if he had no information that would undermine his belief. The example of Mr. Truetemp illustrates this perfectly. He has no evidence that his thoughts about the temperature are incorrect. Had he taken time to obtain and consider evidence, he would have discovered that his thoughts about the temperature were correct, but he did not obtain or consider any evidence concerning the matter. As a result, he does not know that his thoughts about the temperature are correct.

Take a more commonplace example. If I read a thermometer at the local gas station, and it says that the temperature is 104 degrees, I do not know simply from reading the thermometer that the temperature is 104 degrees. I may not have any evidence that it is untrustworthy, but the objection that gas station thermometers are often inaccurate is not one I can answer or

neutralize, at least not without inquiring about the thermometer. Whether or not the belief-forming process is reliable, which perhaps it is, I do not know whether the information about the temperature is trustworthy or not. Indeed, I may have no view on the matter. I may believe what I see out of habit, but this is not knowledge. This is a central problem for externalism, to wit, that ignorance of our reliability or of other external relationships leaves us ignorant of whether our information is trustworthy. Trust sharpens the epistemic blade.

The Invincibility Objection

There is another objection to historical reliabilism that leads to an important lesson. The objection raised by Cohen is that if we are deceived in such a way that we are invincibly ignorant of the deception, we are justified in what we believe, nonetheless.[19] Cohen's example was the Cartesian demon who deceives us in all our perceptual beliefs. The details of the deception may vary, but let us suppose that the demon clouds our senses and supplies us with deceptive sensory data leading us to believe that we perceive the world, although we actually perceive nothing at all. Since our perceptual beliefs are virtually all erroneous, the process that produces them is not reliable. Yet, Cohen suggests, we are certainly justified in our beliefs. We may have done the best we could to ensure that we were not deceived, attended to what we observe with the greatest circumspection, and noticed no error. Having done the best we could, indeed, the best anyone could do, we are certainly justified in believing what we do.

The intuition is reinforced by noting the difference between two people—one who examines his sensory data with the sort of care that would keep him virtually free from error in normal circumstances and one who forms perceptual beliefs so casually that he would frequently err under the best of circumstances. The former puts together all his information and concludes that he is seeing the path of an alpha particle in a cloud chamber. The other believes this because some person, whom he knows to be scientifically ignorant, has told him that this is what he is seeing. We would wish to say that the former but not the latter was justified in believing that he sees the path of an alpha particle in a cloud chamber, even though both beliefs are produced by processes that are unreliable, given the interventions of the demon.

Reliabilism might be modified to meet the objection, and Goldman has suggested more than one way.[20] The example shows that it is internal factors, not external ones, that make us justified and explain the difference between the circumspect and casual observers above. The sort of justification appealed to in the example is personal justification as explicated in the last chapter. The circumspect observer wins the justification game round arising when the skeptic claims that casual observations are often in error by reply-

ing that his observation is circumspect and not casual. The casual observer loses that round to the skeptic.

The Absentminded Demon

Reliable belief formation does not seem to be a necessary condition of justified belief. There is, however, an important lesson to be learned from reliabilism: the sort of justification required for knowledge is not entirely an internal matter, either. On the contrary, the needed form of justification depends on the appropriate match between what one accepts about how one is related to the world and what is actually the case. To see this, consider a minor amendment in the preceding example in which the demon, in a moment of cosmic absentmindedness, forgets for a moment to cloud our senses, with the result that we really perceive what we think we do. If this moment is one that occurs very briefly as we suddenly awake from sleep and is immediately followed by further slumber to conceal the demonic lapse, we might believe we perceive what, in this instance, we actually do perceive. I might perceive my hand for the first time and believe I see a hand, only to lose consciousness after this formidable event. Do I know that I see a hand in that brief moment? I believe I do, but, since such beliefs are almost all false, I am almost totally unreliable in such matters as is everybody else, though accepting myself to be reliable, I lack knowledge.

I am as much deceived about my being reliable in this case as I would be when confronted with a convincing liar who tells me almost all falsehoods about a party he attended except for one fact which, in a moment of absentmindedness, he accurately conveyed, namely, that he arrived before the host. If I accept all that he tells me and also that he is a trustworthy source of information about the event, I may be personally justified in accepting all that he says, but I do not know that the one truth he has conveyed is a truth. I do not know that he arrived before the host, since my assumption that my informant is trustworthy is in error, even if he has told me the truth in this one instance, and this error is sufficient to deprive me of the sort of justification I require for knowledge. This is the truth about justification contained in reliabilism.

Undefeated Justification and Reliabilism

The account that we have offered of undefeated justification in the last chapter is sufficient to deal with the sort of problem we have just considered. To be personally justified in accepting what another says, one must accept that the person is trustworthy, for, otherwise, the critic can win the justification game by claiming that informants are sometimes untrustworthy or, more directly, that the informant from whom I received the information is an untrustworthy informant. Thus, to be personally justified, I

must accept that the informant is trustworthy. Since that is false, however, I will not be justified in accepting that my informant arrived before the host on the basis of my ultrasystem, which excludes the use of false acceptances to meet objections. I will not be justified on the basis of my ultrasystem, and so I will not be irrefutably justified even though I am personally justified.

Trustworthiness and Reliability

Should we say that a person is worthy of her own trust in what she accepts when she is invincibly deceived, by the evil demon, for example? We may proceed in a way that makes us worthy of our own trust and yet be deceived, for we are fallible creatures who seek truth without any impenetrable armor against deception. We have our defenses, however, which make us worthy of our trust even though we are vulnerable to deception. In the case in which the deception is global, as in the case of the demon, rather than some odd case of local deception, a person may be worthy of her own trust in what she accepts about what she perceives, even though she is invincibly and globally deceived about what she perceives. She may be worthy of her own trust because she is as trustworthy in what she accepts as the circumstances allow. She is circumspect and seeks to detect every error in a way that makes her worthy of her trust before accepting what she does. There is no fault in her; the defect is external to her and lies in the circumstances that invincibly deceive her.

Trustworthiness and Truth

How are we to describe the general problem in what a person accepts if we agree that she is trustworthy in what she accepts? Obviously, the problem is that what she accepts about what she perceives will be globally in error. To be worthy of one's own trust, it suffices that one proceeds in the right way, in a way that one ought to proceed; but, being fallible, one proceeding in these ways is vulnerable to global deception.[21] However, a person who is trustworthy in what she accepts but who is globally in error cannot be relied on to provide us with the truth. She is not a reliable provider of truth, even though she is trustworthy in the way she proceeds to accept what she does.

Thus, we may say that a person may be trustworthy in what she accepts as a result of proceeding in ways that are worthy of her trust even though she is not reliable in providing herself or others with truth in what she accepts. In short, a person may be trustworthy because of her methods but unreliable because of the circumstances. Trustworthiness is a matter of proceeding in a way that makes her worthy of her trust, whereas reliability is a matter of a high frequency of success in obtaining truth in the way she pro-

ceeds. It is a sad but important truth that one may be trustworthy without being reliable, that is, successfully reliable, in the quest for truth.

Nevertheless, the objective at which a person aims in proceeding in a way that makes her worthy of her own trust is to succeed in accepting what is true. She prefers to be trustworthy in what she accepts in order to obtain truth. Moreover, since that is her preference and aim, she proceeds in the ways she does, ways that make her worthy of her trust, because she accepts that these ways of proceeding are reliable ways to succeed in obtaining truth and avoiding error. She accepts, therefore, that when she is trustworthy, she will be reliable. She accepts that if she is trustworthy, then she will be reliable, although, conceding she is fallible, the connection between trustworthiness and reliability must be contingent.

The accepted connection between trustworthiness and reliability is the truth connection. It connects the evaluation system on which the trustworthiness of a person is based at a point in time with the ultrasystem in which error is deleted. If a person is personally justified in accepting what she does, then she accepts, or at least prefers to accept, that if she is trustworthy in what she accepts, then she is reliable in what she accepts. For it is an objection to what she accepts that, though she accepts it in a way that is worthy of her trust, she is, perhaps through no fault of her own, unreliable in obtaining truth and avoiding error in what she accepts. To answer the objection she must accept or prefer to accept that if she is trustworthy in what she accepts, then she is reliable in what she accepts.

Trustworthiness and Reliability in the Justification Game

Put in terms of the justification game based on the evaluation system of a person, a critic can first object that a person is not trustworthy in what she accepts. That objection might be answered on the basis of the evaluation system of the person. She might reply that it is more reasonable for her to accept that she is trustworthy than that she is not because she accepts or prefers to accept that she is trustworthy.

But in a second round of the game the critic makes a second objection, namely, that even if the person is trustworthy in what she accepts, she is not reliable in what she accepts because trustworthiness does not yield reliability. To win this round the person must accept that if she is trustworthy, then she is reliable, or at least she must prefer accepting this to accepting the negation of it. For then she can reply that it is more reasonable for her to accept than deny that if she is trustworthy, then she is reliable in what she accepts. The objection of the critic is that trustworthiness and reliability are unconnected. This objection must be answered on the basis of the ultrasystem of the person.

Conversion to Irrefutable Justification

Thus, the conversion of personal justification to irrefutable or undefeated justification requires the truth of the following principle:

(TR) If I am trustworthy in what I accept, then I am reliable in obtaining truth and avoiding error in what I accept.

So personal justification requires acceptance of (TR) or preference for accepting (TR) over the negation of it. Otherwise, the critic will win the justification game. Personal justification is, of course, necessary for knowledge, but such justification, to be converted to knowledge, must be irrefutable or undefeated. For the justification to be undefeated, acceptance of (TR), or preference for accepting (TR) over the negation of it, must be contained in the ultrasystem. The reason is easy to appreciate. If (TR) is false, then the ultracritic in the ultra justification game can appeal to the falsity of (TR) to win a round in that game and the justification is defeated. To put it another way, the falsity of (TR) will serve to defeat or refute the personal justification of the person. So the truth of it is required for the conversion of personal justification to knowledge.

Internal trustworthiness must match external reliability to obtain irrefutable and undefeated justification. Hence the account offered above incorporates the reliabilist insight and explains how we fail to obtain knowledge when testimony or our source of information is untrustworthy or unreliable. We require that our testimonial source of information be trustworthy to answer the objection that he is not and, beyond that, we require of the other as we do of ourselves, that the trustworthiness of the other result in reliability in what he accepts. For if he is trustworthy but unreliable, perhaps because he is deceived through no fault of his own, the objection that this is so will be unanswered.

The appeal of reliabilism and the other forms of externalism may, moreover, be easily understood in terms of the coherence theory and the account of undefeated justification contained therein. To oversimplify a bit, personal justification depends on our background information about the relationship of acceptance to the truth frequency of what is accepted—about nomological or statistical correlations, about counterfactual dependence, or about reliable processes. This information is contained in my acceptance system. I know that I see my cat sitting on papers on the desk. I accept that I would not believe that I see a cat if it were not true that I see him. I accept that my believing I see a cat is correlated with my seeing a cat, though I would not put it that way. I accept that always, or almost always, I see a cat when I think I see one because my accepting that I see a cat results from a trustworthy process or way of proceeding

that is also reliable. It is my acceptance of these things that converts merely accepting that I see a cat into personal justification and victory in the justification game.

For that victory to be converted into irrefutable and undefeated justification, however, what I accept about these things or prefer to accept must also be true. The conversion of mere acceptance into personal justification depends on my accepting the things about myself that render them transparent. By contrast, the externalist mistakenly assumes that the mere existence of the external relations and the reliability of them, however unnoticed and even invincibly opaque to the person, would be sufficient to convert true belief into knowledge. The conversion also depends on the truth of these things I accept about myself and my reliability. The error of externalism is to fail to notice that the subject of knowledge must accept and be correct in accepting that the internal conditions of trustworthiness are connected with the externalist relations. The insight of externalism is the claim that the external conditions of reliability of our trustworthiness must, indeed, hold true.

Causation and Justification: The Basing Relation

The truth contained in reliabilism is, however, concealed by an error. What a person originally believes as a result of prejudice may later be accepted on the basis of scientific evidence. Therefore, the reliabilist must be in error when he claims that it is what *originates* a belief that converts it into a justified belief and knowledge. This is, in effect, to confuse the *reason* a person has for believing something with the *cause* of his believing it. The confusion is such a common one that we might name it the *causal fallacy.* It is easy to see how the fallacy arises. When a person's justification for her belief is based on evidence, then she believes what she does *because* of the evidence. This suggests a causal account of what is involved when the justification of a belief is based on evidence. It suggests that the notion of a justification being based on evidence should be explicated in causal terms. Following this proposal, a person's justification for her belief is based on certain evidence if and only if her belief is causally related in some specified way to the evidence. How to specify the exact way in which the belief must be causally related to the evidence would remain a problem on this approach, but it would be a problem of detail rather than of principle. All such theories must be rejected, however.

Often the evidence on which a justification is based does causally explain the existence of the belief, and it may even be admitted that sometimes the belief is justified because of the way in which it is causally explained by the evidence. Nevertheless, it is also possible for a justified

belief to be causally independent of the evidence that justifies it. Indeed, it may well be that the evidence in no way explains why the person holds the belief, even though her justification for the belief is based on the evidence. The evidence that justifies a person's belief may be evidence she acquired because she already held the belief, rather than the other way round. This is to be expected, since it is common sense to distinguish between the reasons that justify a belief and the causes that produce it. The causes of belief are various and, though the reasons we have for a belief sometimes cause the belief to arise, the belief may also arise from some other cause than having the reasons that justify it. Having the reasons we do may justify the belief, however, even though they have no causal influence on the belief at all.

An example will illustrate. It is easy to imagine the case of someone who comes to believe something for the wrong reason and consequently cannot be said to be justified in his belief; but as a result of his belief, he uncovers some evidence that completely justifies his belief. Suppose that a man, Mr. Raco, is racially prejudiced and as a result believes that the members of some race are susceptible to some disease to which members of his race are not susceptible. This belief, we may imagine, is an unshakable conviction. It is so strong a conviction that no evidence to the contrary would weaken his prejudiced conviction, and no evidence in favor would strengthen it. Now imagine that Mr. Raco becomes a doctor and begins to study the disease in question. Imagine that he reads all that is known about the disease and discovers that the evidence, which is quite conclusive, confirms his conviction. The scientific evidence shows that only members of the race in question are susceptible to the disease. We may imagine as well that Mr. Raco has become a medical expert perfectly capable of understanding the canons of scientific evidence, though unfortunately he becomes no less prejudiced in his beliefs as a result of this.

Nevertheless, Mr. Raco understands and appreciates the evidence as well as any medical expert and as a result has reason for accepting his belief that justifies his acceptance of it. He has discovered that his conviction is confirmed by the scientific evidence. He knows that only members of the other race are susceptible to the disease in question. Yet the reasons that justify him in this acceptance of his belief do not causally explain the belief. The belief is the result of prejudice, not reason, but his present acceptance of it is confirmed by reason that provides the justification for accepting the belief. Prejudice gives Mr. Raco conviction, but reason gives him justification for acceptance.

Harman and others, most notably Marshall Swain and Alvin Goldman, have suggested that a belief is based on evidence only if the evidence conditionally or partially explains the belief.[22] The idea is that the belief must be

causally sustained by the evidence, even if it is not originated by the evidence on which it is based. Again, in the typical case, this will be true. Usually the reasons a person has for a belief can be expected to have some causal influence on the belief, even if they do not originate that belief. However, there is no better reason for supposing that the evidence which justifies accepting a belief must partially explain or causally sustain the belief than for supposing that it must originate it. We may suppose that the evidence justifying Mr. Raco's acceptance of his belief does not in any way explain or causally sustain his belief. What explains and sustains his belief is his prejudice. His belief is neither strengthened nor explained by his discovering the evidence for it. His prejudice gives him the strongest level of conviction, and the evidence adds nothing to the strength of it.

One might, however, suggest that his conviction is conditionally or counterfactually explained or sustained by the evidence, nonetheless. It might be proposed that if Mr. Raco were not to believe what he does out of prejudice, he would believe it as a result of the evidence. This is again likely, but it need not be so. Imagine that Mr. Raco is so dependent on his prejudice that if he were to cease believing what he believes out of prejudice, his racist worldview would collapse, leaving him dysfunctional and uninfluenced by reason. To avoid such an objection one might propose, as Swain did, that to say the belief is sustained by the evidence is only to say that if Mr. Raco were not to believe what he does out of prejudice but were to continue to believe it nonetheless, then he would believe it as a result of the evidence. Perhaps this is to be expected, but must it be so? Again suppose that were Mr. Raco to cease to believe what he does out of prejudice, he would become quite mad and, uninfluenced by reason, he would believe what he does as a result of madness.

The point is the one with which we began. Though evidence ordinarily has some influence over belief or would have if other factors were to lose their influence, this is really incidental to justification of acceptance. The analogy between justification and validity explains why. If a person validly deduces a conclusion from something he knows, this may cause him to believe the conclusion or may influence his belief in the conclusion. But the validity of the inference does not depend on this causal influence. If valid deduction had no influence whatever on whether a person believed the conclusion, that would not undermine the validity of the inference. Similarly, if someone justifies some conclusion on the basis of something he knows, this may cause him to believe the conclusion or influence his belief in the conclusion. The justification of his conclusion, however, does not depend on the causal influence. Thus, a person may justify a second belief in terms of a first belief and the justification of the second belief may be based on the first without the second belief being causally influenced thereby.

Acceptance, Belief, and Justification

The preceding discussion rests on a distinction between explaining why a person believes something, on the one hand, and explaining how he knows it as a result of accepting it, on the other. When a person knows that his belief is true, he accepts it. Usually a person accepts what he believes, though not always, for acceptance involves evaluation for the purpose of obtaining truth and avoiding error, and sometimes we evaluate what we believe negatively in terms of these goals. In other cases, belief, though serving other purposes, may be consistently accepted in terms of the truth objective as well. The explanation of why a person believes what he does may have something to do with his having the evidence he does, but it need not. His acceptance, which goes beyond the mere belief and includes the evaluation of it, is justified by the evidence. The explanation for the belief may rest on political, erotic, or other extraneous influences, but the explanation of how a person knows that his belief is true, when the justification of his acceptance of the belief is based on evidence, must be in terms of the evidence. It is how a person knows what is explained by evidence that justifies his acceptance of it. Why he believes what he does may be explained by almost anything. Justification for acceptance of a belief that is known to be true is based on specific evidence if and only if his having that evidence explains how he knows that the belief he accepts is true. The evidence explains how the person knows, moreover, if and only if the evidence justifies the person's acceptance of the belief. The manner in which evidence justifies acceptance of a belief is explained in the account of personal and undefeated justification. Evidence that justifies acceptance of a belief in a way that converts justification to knowledge is evidence that provides irrefutable justification for acceptance within the ultrasystem of the person, from which all error is deleted.

The idea of evidence explaining how a person knows may be further clarified by recalling once again that our primary concern is to provide a theory to explain how people know that the information that they possess is correct. If the evidence that a person has justifies her acceptance of the belief that *p*, then the evidence explains how she knows that the information that *p* is correct. Similarly, if a person is asked how she knows that *p*, her reply will be to justify the claim that *p* in terms of her evidence. It is her evidence that shows that she knows and how she knows. Thus, a justification based on evidence explains how a person knows that *p* if that justification would be a correct answer to the question, How do you know that P?

Reliability and the Justification Game

Reliability enters into justification not by originating belief but by supporting acceptance of the belief in the justification game. It is this game that re-

veals the role of the evaluation system in justified acceptance. Consider the justification game played by the prejudiced man, Mr. Raco, *before* obtaining the scientific information.

Claimant: The members of that race suffer a disease to which members of other races are not susceptible.
Skeptic: You believe what you do as the result of prejudice.
Claimant: It is more reasonable for me to accept that I do not believe what I do as a result of prejudice than to accept that I believe what I do as a result of prejudice. (I am quite unprejudiced concerning members of the race in question; it is just that they are inferior.)

This personal justification would fail to convert into irrefutable and undefeated justification. The claimant's error concerning his prejudice would disqualify the last move above in the ultra justification game.

After acquiring the scientific information, the claimant is in a position to neutralize the claim of the skeptic in the justification game and defend his acceptance of the belief by making the following reply to the claim of the skeptic above:

Claimant: Though I believe what I do out of prejudice, I now accept what I do on the best scientific evidence. It is as reasonable for me to accept *both* that I believe what I do out of prejudice *and* that the best scientific evidence shows that what I now accept on this evidence is true as to accept that I believe what I do out of prejudice alone. (In the standard medical reference work concerning this disease, it is stated that only members of the race in question are susceptible to the disease. This has been confirmed by recent studies, which I have cited.)

This move succeeds in the ultra justification game. The claimant wins the round as a result of neutralizing the objection, and his move cannot be disqualified. Whatever his moral failings, as a result of obtaining scientific understanding, he is victorious in the ultra justification game. He is, therefore, personally justified and also irrefutably justified in accepting what he does because the scientific evidence is correct.

The preceding reflections illustrate the point that the evidence which justifies a person in accepting something must explain how the person knows that p rather than why he believes it. The scientific evidence explains how the person knows by explaining how he can be victorious in the justification game. Usually, what makes a person victorious in the justification game is closely connected to what makes him believe what he does. But the connection is not essential to justification of acceptance. As a result, the reliability essential to justification of acceptance is not the re-

liability of the process that produces or causally sustains belief. What is essential is that the trustworthiness of the evidence is connected with a high frequency of success in obtaining truth and avoiding error in what we accept. The trustworthiness of the evidence makes us trustworthy in what we accept in a way that is connected with reliability, though we are fallible and subject to such human frailties as prejudice. It is such trustworthiness of acceptance and not infallibility that is needed for justification. In epistemology as in life generally, you do not have to be perfect in order to be justified.

Trustworthiness and Prejudice: An Objection

Let us reflect on an objection to the preceding line of thought. It is as follows. It has been affirmed that a person must be trustworthy in order to be personally justified. It has also been affirmed that trustworthiness is a matter of how one changes in the face of objections to what one believes that one cannot meet. A person who does not change what he believes in line with the evidence is not trustworthy or reliable in what he believes. Raco does not change what he believes in line with the evidence because his prejudice renders him invulnerable to evidence that is contrary to what he believes out of prejudice. Therefore, he is not trustworthy in what he believes, and his belief, though accepted on scientific evidence, is not justified. So runs the objection that prejudice destroys trustworthiness, reliability, and justification.[23] What is the answer to this objection?

The answer to the foregoing objection is that Raco, though he is not trustworthy in the way he forms and changes beliefs, may be trustworthy nevertheless in how he evaluates what he believes and what he accepts. We noted that acceptance and belief may come into conflict and that the system of acceptance and the system of belief may pull apart in extreme cases. We find in the case of Raco a person whose belief system and acceptance system are more radically separate than they are for the rest of us. He may be trustworthy in what he accepts but not in what he believes. He may be trustworthy in how he changes what he accepts but not in how he changes what he believes. He may, in short, be trustworthy in how he changes concerning acceptance but not with respect to belief.

Must acceptance be causally explained by the evidence contained in the evaluation system in order to be justified by it? Of course, acceptance might be *causally* influenced by the evaluation system of the person, that is, by the acceptances, preferences, and reasonings of the person that justify the acceptance. Once again, however, this does not seem necessary. The evaluation system of the person provides justification for what a person accepts by providing answers to some objections and neutralization of others. If we

study the underlying causal processes, we may find that the evaluation system is a causal factor in the explanation of responses of the subject to critical objections. But what is essential is that the person understands how to meet and neutralize the objections to what he accepts.

It is, therefore, the capacity to respond to objections with understanding rather than actually responding to them that is crucial to justification. Is there some causal account to be given of how a person acquires such understanding? No doubt there is, just as there is a causal account to be given of how a person acquires understanding of how to reason validly. It would be a mistake, however, to conclude from this that the validity of reasoning to a conclusion on the basis of premises is to be explained by the causal account of how we understand it or acquire understanding of it. It is the same mistake to conclude from the fact that there is a causal account of how we acquire understanding of how to justify acceptance (by answering or neutralizing objections on the basis of our evaluation system) that, therefore, justification for acceptance on the basis of our evaluation system is to be explained by the causal account of how we understand it or acquire understanding of it. This is the causal fallacy.

We reason from premises to conclusions, and this has a causal explanation. Nonetheless, the validity of the reasoning is not to be explained causally but in terms of a relation between the premises and what is concluded. We reason from our evaluation system to defend what we accept, and this has a causal explanation. The justification of this defense is not explained *causally* in terms of the relation between the system and what is accepted. That relation is not causal. Validity and justification are properties based on the principles of logic and evidence. The relations of logical consequence and evidential support are not causal relations, as Chisholm and Feldman have shown us.[24]

Externalism, Foundationalism, and Coherence: An Ecumenical Reconsideration

The foregoing articulation of the coherence theory of justification suggests that there is some merit in the foundation theory and in externalism, which we have preserved in our theory. It is, therefore, time to turn from criticism to ecumenicalism. The foundation theory held some introspective, perceptual, and memory beliefs to be self-justified. We argued that the justification of the acceptance of all such beliefs depends on background information concerning our trustworthiness in such matters. Thus, it is coherence with such information in our acceptance system that produces the justification. Nevertheless, we concede that the acceptance of some beliefs is justified without inference because we accept ourselves to be trustworthy in such

matters. A principle of our trustworthiness is needed to convert mere acceptance into personally justified acceptance.

Moreover, though the principle of our trustworthiness must cohere with what we accept about our successes and failures in past epistemic employments, the principle of our own trustworthiness contributes to its own personal justification. Part, but not all, of what makes us personally justified in accepting that we are trustworthy is that we do accept that we are. If we did not accept that we were trustworthy, there would be an unbeatable skeptical challenge to any claim we made in the justification game, to wit, that we are untrustworthy in what we accept. To answer that objection, we must accept that we are trustworthy. So, there appears to be at least one thing that we accept, one important and fundamental thing, that is in part self-justified, as the foundationalist contended, even if it is not those introspective, perceptual, and memory beliefs that he most favors. To be personally justified one must accept some principle of trustworthiness that is in part self-justified.

To be personally and irrefutably justified as well, some principle of trustworthiness we accept must be true. Otherwise, the skeptical challenge that we are not trustworthy in what we accept would not be answered in the ultra justification game. The insight of externalism is the contention that there must be some truth connection between our trustworthiness in accepting what we do and the truth of what we accept. We accept both that we are trustworthy in what we accept and that if we are trustworthy in what we accept, then we shall be reliably successful in accepting what is true and avoiding accepting what is false. The correctness of what we thus accept about our trustworthiness and how it is related to reliability yields the truth connection required for irrefutable and undefeated justification.

Externalism is motivated by doubt about whether what we accept can supply the truth connection. The reason for the doubt is the assumption that it is psychologically unrealistic to suppose that beliefs about our beliefs are necessary for knowledge. Such higher-order beliefs about beliefs are not, of course, necessary for receiving and relaying information. Even a thermometer is capable of that. Acceptance of our trustworthiness and reliability is, however, necessary for knowledge. Is it unrealistic to suppose that people accept that they are trustworthy? Some unrealistic theory of belief maintaining that all beliefs are occurrent states may yield the consequence that we lack such beliefs, but our theory of acceptance explains how we accept that we are trustworthy. The mental state of acceptance is a functional state, one that plays a role in thought, inference, and action. We think, infer, and act in a way manifesting our trust in what we accept.

Thus, it is appropriate and not at all unrealistic to suppose that, in addition to the other things we accept, we accept our own trustworthiness and the reliability of it as well. We have provided the truth that supplies the

truth connection required by the externalist in the form of a justified principle of our own trustworthiness and a connection to reliability. We cannot be accused of chauvinism in claiming that irrefutable justification is the result of coherence with an acceptance system incorporating the principle of our own trustworthiness and the truth of what we accept. Unless we are trustworthy in what we accept, neither we nor our adversaries can be justified in what we accept and we must all concede the day to the skeptic. If we are trustworthy in the quest for truth and trustworthiness is successful, as we accept it is, then a target acceptance may cohere with our evaluation system and our ultrasystem to yield undefeated justification and knowledge. The attainment of knowledge, like so many other benefits in life, rests on self-trust and the success of it.

Introduction to the Literature

The most important defenders of externalism and reliabilism are Alvin Goldman, *Epistemology and Cognition;* D. M. Armstrong, *Belief, Truth, and Knowledge;* Fred Dretske, *Knowledge and the Flow of Information;* and Robert Nozick, *Philosophical Explanations,* chapter 3. A refined attempt to combine reliabilism with a theory of justification is contained in *Reasons and Knowledge,* by Marshall Swain. For an important critical account of reliabilism, see Stewart Cohen's "Justification and Truth," as well as Richard Feldman's "Reliability and Justification." The original and classic article on naturalized epistemology is W. V. O. Quine's "Epistemology Naturalized," in his *Ontological Relativity and Other Essays.*

Notes

1. Alvin Goldman, "A Causal Theory of Knowing," *Journal of Philosophy* 64 (1967): 357–72.

2. Alvin Goldman, *Epistemology and Cognition* (Cambridge: Harvard University Press, 1986).

3. D. M. Armstrong, *Belief, Truth, and Knowledge* (Cambridge: Cambridge University Press, 1973); Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981).

4. Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981), chap. 3.

5. Alvin Goldman, "What Is Justified Belief?" in *Justification and Knowledge,* ed. G. S. Pappas (Dordrecht: Reidel, 1979); see also Goldman, *Epistemology and Cognition.*

6. David Hume, *A Treatise of Human Nature* (Indianapolis: Bobbs-Merrill, 1977).

7. W. V. O. Quine, "Epistemology Naturalized," in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).

8. D. M. Armstrong, *Belief, Truth, and Knowledge;* Fred Dretske, *Seeing and Knowing* (London: Routledge & Kegan Paul, 1969).

9. The "braino" example comes from James M. Cornman, Keith Lehrer, and G. S. Pappas, eds., *Philosophical Problems and Arguments: An Introduction,* 3d ed. (New York: Hackett, 1987), 54–55.

10. Dretske and Goldman have both discussed replies to skepticism that distinguish relevant from irrelevant alternatives. See Goldman, "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73 (1976): 771–91; and Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971): 1–22. See also Dretske, *Knowledge and the Flow of Information.*

11. Goldman, "Causal Theory of Knowing."

12. Dretske, *Knowledge and the Flow of Information.*

13. Alvin Goldman, *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge: MIT Press, 1992).

14. Nozick, *Philosophical Explanations.*

15. Alvin Plantinga, *Warrant and Proper Function* (New York: Oxford University Press, 1993), 194.

16. D. M. Armstrong, *Belief, Truth, and Knowledge.*

17. Nozick, *Philosophical Explanations;* Dretske, *Knowledge and the Flow of Information.*

18. Goldman, *Epistemology and Cognition,* 62–63, 111–12.

19. Stewart Cohen, "Justification and Truth," *Philosophical Studies* 46 (1984): 279–96; see also Keith Lehrer and Stewart Cohen, "Justification, Truth, and Coherence," *Synthese* 55 (1983): 191–207.

20. The most influential modification, one later rejected by Goldman, involves a restriction to normal worlds. See Alvin Goldman, *Epistemology and Cognition,* 107–9; and "Strong and Weak Justification," in *Philosophical Perspectives,* vol. 2, ed. J. Tomberlin (Atascadero, Calif.: Ridgeview, 1988), 51–70.

21. David Truncellito explored the need to proceed in the right way in his doctoral dissertation, and I am indebted to him for this construal. For a critical discussion of deontological epistemology, the epistemology of what we ought to believe, see Alvin Plantinga, *Warrant: The Current Debate* (New York: Oxford University Press, 1993), 19–24.

22. See Gilbert Harman, "Knowledge, Reasons, and Causes," *Journal of Philosophy* 67 (1970): 841–55; also his *Change in View* (Cambridge: MIT Press, 1986); and *Thought* (Princeton: Princeton University Press, 1973); as well as Marshall Swain, *Reason and Knowledge* (Ithaca, N.Y.: Cornell University Press, 1981). For an argument that no revision of Swain's account can succeed, see Jonathan Kvanvig, "Swain on the Basing Relation," *Analysis* 45 (1985): 153–58.

23. The objection is due to Rachael Poulsen.

24. See R. M. Chisholm, *Theory of Knowledge,* 3d ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1989); and Richard Feldman, "Reliability and Justification," *Monist* 68 (1985).

9

Skepticism, Virtue, and Context

WE SAY WE KNOW, but do we? Skeptics have denied it and they have had an influential history. We shall, in the light of our epistemology, assess the genuine merits of skepticism. We have mentioned the skeptic from time to time and have used the skeptic as a heuristic opponent. Now we turn to the philosophical skeptic who genuinely challenges our claim to knowledge. We shall consider whether there is any answer to the skeptic to be found in an appeal to intellectual virtue and the context in which we claim to know what we do. We shall find that intellectual virtue and trustworthiness manifest the same disposition, and thus that virtue epistemology, defended by Ernest Sosa, Alvin Goldman, Linda Zagzebski, John Greco, and others, is relevant to the problem of skepticism.[1]

The question of whether intellectual virtue, or trustworthiness, succeeds in attaining its goal depends on the circumstances and in this way, as Michael Williams suggests, knowledge is contextual.[2] However, other attempts to deal with the problems of skepticism by appeal to the context in which we attribute knowledge, such as that of Keith DeRose, for example, do not deal with the skeptic in a satisfactory manner.[3] We shall find that the skeptic can be answered, including the skeptic who suggests the most ingenious hypotheses of deception, but our answer should not be taken as a proof that the skeptic is wrong. He cannot be proven wrong. Nevertheless, we may know what he denies that we know, including that we know that we know.

Skepticism and Agnoiology

Skepticism comes in different depths. Shallow forms deny that we know a few of the things we claim to know, and the deepest form denies that we

know anything at all.[4] Deeper forms of skepticism are based on the ubiquitous chance for error. Plain people, who comfort themselves in the snug foothills of accepted opinion, overlook the possibilities for error residing in our most familiar beliefs. In the mind of the dogmatic, what is familiar comes, through long acquaintance, to appear completely dependable and wins unquestioning confidence. The philosophical skeptic, inclined to question when others are drawn to dogmatic tranquillity, discovers the risk of error in our most trusted convictions. On this discovery, she constructs an *agnoiology,* a theory of ignorance.

Of course, skeptics who have denied that we know what we say we do have frequently been moved by more than a passion for the study of agnoiology. Often they espouse some theory that conflicts with common opinion. Skepticism is defended to win consideration for their own theories. In reply, commonsense philosophers, like Thomas Reid and G. E. Moore, have rejected such speculative theories on the sole grounds that they conflict with common sense.[5] The beliefs of common sense are innocent, they say, until proven guilty and constitute knowledge unless they are shown not to. Skeptics have been accused of semantic deviation, logical absurdity, and triviality. In an earlier chapter, we argued that what the skeptic says is semantically acceptable, logically consistent, and highly contentious. Rather than attempt to dismiss her abruptly by some superficial artifice, let us consider what sustains her argument.

There are a number of classical skeptical arguments appealing to dreams and hallucinations purporting to show that, whatever we take to be true, there remains some chance of error.[6] However, skeptical argumentation does not depend on these appeals. They are simply familiar ways of explaining how people might err. It matters little what the source of error may be. What is critical is most obvious. People often accept what is false and, when what they accept is true, there is some chance that they might have erred. This is the fundamental skeptical premise.

Conception and the Chance of Error

There are a variety of ways in which a skeptic may press this premise. Such arguments have the merit of calling our attention to some possibility of error we overlook. For example, a skeptic may base his argument on the nature of human conception. Experience by itself tells us nothing. Knowledge requires the application of concepts and background information to experience. The best entrenched concept remains constantly subject to total rejection. In the pursuit of truth, we may discard any concept as lacking a denotation. Any concept may be thrown onto the junk heap of repudiated concepts along with demons, entelechies, and the like. Moreover, any discarded concept can be refurbished. Because the concepts we reject may be

better than the ones that supplant them, we may have to recycle what we discard. No concept or belief is sacrosanct in the quest for truth, and there is always some chance that any one may be cast off as misleading and erroneous.

The foregoing remarks describe more than a mere logical possibility. It is not only logically possible that any belief is in error, but there is some genuine chance that it is so. The beliefs that have been most cherished and in which people have placed their greatest confidence, for example, the belief in witches, have been demoted from literal truths to figures of speech. Strictly speaking, there is no such thing. The concept of a witch, aside from use as a figure of speech, is a relic of religious conceptualization that is no longer tenable in an impartial and disinterested search for truth. This merely illustrates how, in the flux of conceptual change and innovation, any concept may be rejected for the sake of conceptual improvement and increased veracity.

We must note in passing that the concept of belief, indeed, even the concept of a concept, is no more secure than any other. Some materialists have said that belief is essentially mental, and consequently that there is no such thing as belief.[7] We cannot consider such materialism and the implications of it for our theory of knowledge here. Such materialism would require that, if we have referred to anything real in the world when speaking incorrectly of belief, what is real may be correctly described using only a materialistic vocabulary.

The skeptic is correct, we concede, in affirming that the chance of error is always genuine. Skeptics have sometimes invented fanciful hypotheses to illustrate that there is some chance of error in our most secure beliefs, those of perception, for example. Let us recall some mentioned earlier. Descartes imagined a powerful demon intent on deceiving us into believing we perceived a material world when we do not by causing us to have sensations imitating those we would expect when we perceive a material world.[8] Hilary Putnam imagined a scientist who removed brains, placed them in a vat, and provided them with electrical stimulation imitating the stimulation of senses resulting in perceptual beliefs.[9] Less radical scientific manipulation has been proposed whereby a device is implanted in the brain, called a braino, which is controlled by a computer in the hands of a scientist and can be used to deceive us into believing we perceive a material world as Descartes' demon and Putnam's scientist would do.[10] These hypotheses present us with invincible deception. There is nothing we or our brains can do to expose the deception. One may, while agreeing to the logical possibility of such hypotheses, be inclined to deny that there is any real chance that they are true. Since the imagined deception is invincible, however, there does not seem to be any way of ruling out the chance, however remote and minute, that the deception is real. We shall return to the question of how

we might answer such hypotheses, but the chance of error in what we accept does not depend on them, however useful they are in illustrating our vulnerability to deception. Our fallibility as we conceptualize what we experience and change how we do it, discarding concepts that formed our discarded beliefs, replacing or recycling old concepts and beliefs with new or recycled ones, is sufficient to yield the conclusion that there is always some chance that we err. The flux of conceptual change must convince us of our fallibility. Thus, we grant the skeptical premise that if *S* accepts that *p*, then there is some chance that *S* is incorrect. Must we also accept the skeptical conclusion that we are ignorant?

An Answer to Skepticism: Fallibility, Not Ignorance

To sustain skepticism, a skeptic must go on to argue that if there is some chance that *S* is incorrect in accepting that *p*, then *S* does not know that *p*. On the analysis of knowledge that we have articulated, this premise is unavailable. It does not follow from the premise that there is some chance that *S* is incorrect in accepting that *p*, that *p* is not true, or that *S* does not accept that *p*, or that *S* is not justified in accepting that *p*, or that *S*'s justification is defeated. Even if *S* accepts that there is some chance that he is incorrect in accepting that *p*, it may, nevertheless, be just as reasonable for him to accept that *p* and at the same time to admit the objection that there is some chance he is in error as to accept merely the objection that there is some chance that he is in error. As a result, the objection will be neutralized which allows for *S* to be justified in accepting that *p* while admitting the objection.

In the interests of obtaining truth, it may be reasonable to accept something one does while also accepting one's fallibility, that is, accepting that there is some chance that one might be in error. The critic in the justification game may always cite the chance of error as an objection in the justification game, but the claimant can also neutralize it. Our fallibility is an insufficient basis for skeptical victory. We may accept the premise of the skeptic concerning conceptual change and the universal chance of error implicit therein without accepting the deep skeptical conclusion of universal ignorance.

With this reply to skepticism set forth, we hasten to note that in some ways our position is very close to that of the skeptic, for very often when people claim to know something, they claim to know for certain. If they do know for certain, then there must be no chance that they are in error. Hence, in agreeing that there is always some chance of error, we are agreeing with the skeptic that nobody ever knows for certain that anything is true. Joining hands with the skeptic in this way will win us little applause

from those dogmatists who never doubt that people know for certain what they claim to know.

Thus, our theory of knowledge is a theory of knowledge without certainty. We agree with the skeptic that if a person claims to know for certain, he does not know whereof he speaks. However, when we claim to know, we make no claim to certainty. We conjecture that to speak in this way is a departure from a customary use of the word 'know.' Commonly, when people say that they know, they imply they know for certain, and they assume that there is no chance of being in error. This assumption enables them to lay aside theoretical doubts and to pretend they proceed on certain grounds. Such a pretense offers comfort and security in practical affairs and often in scientific investigation as well. Nonetheless, it is a pretense exposed by the skeptic and repudiated by those who seek the truth. We, like the skeptic, deny that our beliefs have any guarantee of truth. We, like the skeptic, admit there is a genuine chance that any of our beliefs may be false. We, like the skeptic, acknowledge that there is some chance, however small and remote, that the skeptical hypotheses which skeptics have conceived to call our dogmatic assumptions into doubt are true and cannot be ruled out by semantic shenanigans or appeal to the fiat of common sense.

Our only reply to the skeptic is that even if there is some chance that any of our beliefs may be in error and even if, therefore, we do not know for certain that any of them are true, still some of the things we accept are things we are justified in accepting because all objections are answered or neutralized on the basis of our evaluation system. Of course, what we accept may be wrong—we are fallible—but if enough of what we accept is correct, then our justification will be undefeated and we will have knowledge. If we are sufficiently correct in what we accept and we can distinguish between when we are trustworthy in what we accept and when we are not, then we may know what we think we do despite the risk of error that we confront. If we were massively mistaken, as we would be if the Cartesian demon were loose in the land, then we would lack knowledge. A merely conceivable demon cannot reduce us to ignorance, however.

Intellectual Virtue and Trustworthiness

The point of the preceding argument can be elucidated by appeal to a notion of virtue in what has become known as virtue epistemology as developed by Sosa, Greco, and Zagzebski, as well as others.[11] Some of those who have been advocates of virtue epistemology have thought of themselves as developing an epistemology that placed virtuous character rather than knowledge at the center of the stage. Nevertheless, the conception of intellectual virtue is much like that of intellectual trustworthiness and plays an important role in the analysis and theory of knowledge.

What is intellectual virtue? It is a virtue that aims at an intellectual goal. What might that be? To obtain truth and avoid error in one's intellectual endeavors, on the present account, to accept what is true and avoid accepting what is false. How might one do this? One must seek to reason validly in deduction and cogently in any argument as an exercise of intellectual virtue. One must consider the objections to the candidates for acceptance and only consider oneself justified in accepting things when one can meet the objections. One must be ready to change what one accepts when the objections cannot be met in order to avoid accepting what is false. In short, therefore, to proceed in an intellectually virtuous manner in what one accepts is to proceed in an intellectually trustworthy manner in what one accepts.

Are intellectual virtue and trustworthiness the same? They are both dispositions to proceed in a manner to achieve the same goal, and so the dispositions with respect to what one accepts and the way one proceeds may be the same in both cases. Thus, the dispositions underlying intellectual virtue and intellectual trustworthiness may be the same. Conceptually, however, they are distinct. There is no contradiction in the idea that a person might be trustworthy and not virtuous or virtuous and not trustworthy. Nevertheless, when the goal of both is the same—to accept what is true and avoid accepting what is false—virtue in the pursuit of that goal will render one trustworthy in the pursuit of the goal, and a person who is trustworthy in pursuit of the goal pursues the goal as virtue requires. A person who would be trustworthy cannot ignore what virtue demands, and one who conforms to those demands will be trustworthy. This equivalence is material, however, and not logical, for it is logically possible that virtue and trustworthiness should divide. If someone who is trustworthy instructs me to accept something that it is not at all intellectually virtuous for me to accept, I may be trustworthy but not virtuous in accepting what the person instructs me to accept.

The test case for this material equivalence of intellectual virtue and trustworthiness, though the notions are conceptually distinct, is the case of invincible deception, like the deception of the demon described by Descartes. Suppose we live in such a world with an active demon and are invincibly deceived in what we accept so that we accept what is false and fail to accept what is true. We may, nonetheless, distinguish the intellectually virtuous person from the intellectually irresponsible one in such a world, and indeed we would proceed to do so in the same manner as at present, since the world would seem exactly like the present world.

Cohen has argued that those who are justified in accepting what they do in the present world, if they proceeded in the same way in the demon world to accept there what they do here, would be justified in what they accept in that world. But if they are justified, then they must be trustworthy in the

demon world, for if they were not, then they would not be justified in what they accept.[12]

The deceived inhabitants of the demon world are as intellectually virtuous in what they accept as they would be in the present world. Are they equally trustworthy? If they proceed in an intellectually virtuous manner, then they are as intellectually trustworthy as the circumstances permit. They are intellectually trustworthy as well as intellectually virtuous because trustworthiness and virtue depend on how one proceeds in pursuit of the goal of accepting what is true and avoiding accepting what is false. Consider those who proceed in the same way we do in our present world but inhabit another world in which they are invincibly deceived. They are as faultless in the way they proceed in pursuit of the goal of truth as we are. They are as intellectually trustworthy or intellectually virtuous in their world as they would be in our world or as we are in our world.

There is, however, an objection that a critic could raise in the justification game which is the same as the one posed by the genuine philosophical skeptic. The objection is this. Though virtue and trustworthiness in what one accepts have truth and the avoidance of error as a purpose, they might fail to achieve the purpose in a reliable manner, that is, in a manner having a high frequency of success. Virtue and trustworthiness that are not reliably successful in achieving their purpose may provide some personal satisfaction, but they cannot provide justified acceptance that converts to knowledge.

The objection might be formulated as follows in the justification game:

Critic (or skeptic): Let us admit that you are intellectually trustworthy and intellectually virtuous, as you claim. You are, nevertheless, in error because such trustworthiness and virtue fail to achieve their purpose. What you accept in this trustworthy and virtuous way is not reliably connected with truth. Reliable connection with truth requires something beyond trustworthiness; it requires a high frequency of success in accepting what is true and avoiding accepting what is false. The demon can deceive you so that you fall into error in all that you accept in an intellectually trustworthy and virtuous way. Your trustworthiness and virtue are useless in the pursuit of their goal to obtain truth and avoid error.

What should the claimant reply? The reply must be that the *critic* or *skeptic* is wrong! To answer the objection, the claimant must, in effect, reply that intellectual trustworthiness and intellectual virtue succeed in achieving their purpose in a reliable manner. So the answer, simply put, is as follows:

Claimant: What I accept in this trustworthy and virtuous way is reliably connected with truth in a successful way. The way of virtue is also the way of truth. It is possible that a demon could deceive us, but we are not so deceived. There is no such demon. It is precisely because there is no such demon that my accepting what I do in a trustworthy and virtuous way enables me to succeed in a reliably successful way in accepting what is true and avoiding accepting what is false. It is more reasonable for me to accept that there is no demon than that there is. It is as reasonable for me to accept both that I am not deceived and that it is possible for the demon to deceive me as to accept merely that it is possible for the demon to deceive me. It is more reasonable for me to accept that trustworthiness and virtue are successfully connected with truth in a reliably successful way than that they are not.

The *claimant* has given a reply in the foregoing comments to the *critic* in the justification game. Is that reply sufficient for the conversion of justification to knowledge? It depends on whether what the things the claimant accepts are true. If they are true, then the claimant will continue to be victorious in the ultra justification game based on the ultrasystem of the claimant as well as in the justification based on the evaluation system of the claimant. Thus, the critical point is that falsity of the skeptical hypotheses of deception is sufficient to convert justification to knowledge. Most critically, it is that justification for accepting that intellectual trustworthiness and virtue must be successfully connected with truth in a reliable manner. We can know that intellectual virtue is successful in its purpose of obtaining truth and avoiding error.

The Duplication Argument: An Objection

The foregoing argument contains an oddity that has troubled many, including the present author. The oddity is that what we would accept as well as what we would experience in the demon world exactly duplicates what we would accept as well as what we would experience in the actual world. So, it seems that if we were invincibly deceived by the demon, we would accept the same reasons for concluding we were not deceived. If we were deceived, then we would accept, as we do in the present world, that we are not deceived. So how can we know that we are not deceived when the reasons we accept for concluding we are not deceived are exactly the same reasons we would accept for concluding we were not deceived if we were deceived?

The reply is that though the *content* of the reasons we accept would be the same, the reasons we accept would not be the same. In the one case, in the actual world, the reasons would be true and in the other case, in the demon world, they would be false. This is a crucial difference. A sound argu-

ment requires premises that are true, not simply premises with a specific content. If we use false premises to reply to objections and justify what we accept, then the resulting justification will be refuted and defeated by our errors. The justification will fail to convert to knowledge. It is precisely because the reasons that we accept to justify what we accept are true that our justification will be irrefutable and undefeated and will convert to knowledge.

It is the very nature of deep deception, as Sosa has lucidly explained,[13] that one accepts that one is not deceived when one is deceived. Assume we know that we are not so deceived. We must know this in spite of the fact that we would believe that we were not deceived even if we were deceived. Thus, we shall not be in position to say that we would not believe we were not deceived if we were deceived. Nevertheless, the justification we have for accepting that we are not deceived may be based on adequate reasons, ones that are true and justify us in accepting that we are not deceived.[14] Their truth ensures that such justification based on an evaluation system of the person will convert to undefeated justification on the basis of the ultrasystem of the person and, thus, constitute knowledge that we are not deceived.

The Merits of Skepticism

Before celebrating victory over the skeptic, however, we should note that the agnoiology of some skeptics is closer to the truth than the epistemology of many dogmatists. We offer no proof that the skeptic is wrong. On our theory of knowledge, whether we win or the skeptic wins the day depends on whether what we accept is correct, and especially on what we accept about when we are trustworthy and when our trustworthiness is successfully connected to truth. We cannot refute the skeptic by appeal to demonstration. We argue against her from our acceptance system, which is precisely what she calls into question.

We may, nonetheless, know that she is wrong. Assuming that our justification for some of the things we accept is sustained by our ultrasystem, we know those things to be true, and indeed we know that we know. This does not mean that we are certain that we know but that we have an undefeated justification for accepting that we know. If we do know that we know, then, of course, we know that the skeptic is mistaken in denying that we know.

We avoid skepticism by constructing a theory of justification without a logical guarantee of truth. On our theory, if people know anything at all, it is because of the correctness of what they accept in their quest for truth. It is what they accept that makes them personally justified in their acceptance, and if enough of what they accept is true, their justification will be unde-

feated and become knowledge. The mere possibility or risk of error is not sufficient to sustain the skeptic. She must deny what we accept, especially that we are trustworthy in a successfully truth-connected way. The skeptic must be correct and we in error to render her victorious. So, whether we know or not depends on whether what we accept about ourselves and our trustworthiness is correct. Surely, that is exactly what we should expect.

To put the matter more precisely, consider the following principle:

T. I am intellectually trustworthy (virtuous) in what I accept in a way that is reliably successful in achieving my purpose of accepting what is true and avoiding accepting what is false.

If *T* is true, the justification a person has for accepting *T* based on accepting *T* would, in normal circumstances, be undefeated. One would expect all objections to *T* to be answered or neutralized on the basis of the ultrasystem of *S* and, therefore, would expect the following equivalent principles to be true:

If *S* accepts that *T* and *T* is true, then *S* is irrefutably justified in accepting that *T*

or

If *S* accepts that *T* and *T* is true, then *S* is justified in accepting that *T* in a way that is undefeated.

Thus, the acceptance of *T,* if *T* is true, may be expected to yield knowledge of the truth of *T.* We may not be able to refute the skeptic who denies the truth of *T* or who advances some skeptical hypotheses implying the falsity of *T.* If, however, we are correct in thinking the skeptic is in error and, in accepting the truth of *T,* then, skeptical machinations notwithstanding, we know that *T* is true and know many other things as a result of this knowledge. We may not have the satisfaction of demonstrating that the skeptic is in error, for the attempt to do so would beg the question. We may, nevertheless, know that the skeptical hypotheses are false and remain content with knowing that the skeptic is wrong even if we cannot prove it.

Skepticism and Closure: An Externalist Caveat

Some antiskeptics, especially externalists such as Fred Dretske and Robert Nozick, have dealt with the skeptic in a different way, which we will now consider. They have claimed that a person knows many things without

knowing that they are not deceived, by the evil demon, for example. They have rejected a *closure principle* as follows: if a person knows at one and the same time that *p* and that if *p*, then *q*, all at once so to speak, then the person knows at the same time that *q*. The closure principle might be formulated with greater precision, of course, but this rough formulation suffices to understand how rejecting such a principle may be useful against a skeptic.[15]

The skeptic advances a skeptical hypothesis, the hypothesis that we are now asleep and dreaming instead of perceiving the external world as we suppose. She goes on to argue that we have no way of knowing that this skeptical hypothesis is false and concludes, therefore, that we do not know that we perceive the things we perceive. The antiskeptic who rejects the closure principle concedes to the skeptic that we do not know her hypothesis is false but denies her inference. He says that we do know that we are perceiving external objects, a piece of paper before us, for example, even though we do not know that we are not now asleep and dreaming. True, he admits, if we are now asleep and dreaming, then we are not now perceiving the piece of paper, but we know that the latter is true even though we do not know that the former is false. Epistemic closure fails, he concludes, and with it the skeptical hypotheses.

The preceding line of thought is typical of externalists because a belief resulting from a reliable process or a belief that tracks truth may have a consequence that does not result from a reliable process or track truth. According to Nozick, I know that *p*, only if my belief that *p* tracks truth so that I would not believe that p if p were not true. However, I cannot claim that I would not now believe that I am not deceived by an evil demon if I were deceived by an evil demon. The reason for this is that if I were deceived by an evil demon, I would believe that I am not deceived by an evil demon, because that is the result of the deception.

Dretske, whose views we considered in the last chapter, has suggested that, though one must be able to exclude relevant alternatives to what one believes in order to have knowledge, the skeptical alternatives fall short of relevance.[16] I do not need to know, and in fact do not know, that I am not asleep and dreaming or that I am not deceived by the demon in order to know that I perceive a table before me, which I do know. Thus, I may know that I see a table even though I do not know that I am not deceived in accepting that I see a table as a result of dreams or demonic deceivers.

The foregoing approach has some appeal. Since we do not think of skeptical hypotheses concerning dreams, hallucinations, Cartesian demons, brains in vats, or the braino as we go about our daily rounds, it is natural to suppose that we do not need to know anything about such matters in order to know the many things we suppose we do, for example, that we perceive external objects. If we know that we perceive those things and do not

know that the skeptical hypotheses are false, then the reason for denying epistemic closure is clear.

The Trilemma of Knowledge and Skepticism

We may consider such antiskeptics as suggesting a way out of a trilemma.[17] Consider the following claims:

1. I know that I see a hand (my hand, in fact) in front of me.
2. If I know that I see a hand in front of me, then I know I am not deceived by a powerful deceiver, for example, the evil demon.
3. I do not know that I am not deceived by an evil demon.

It is clear that all the claims in the argument have some plausibility, and equally clear that not all the claims in the argument can be true when the word 'know' is used unambiguously throughout the argument. The first premise seems obvious. The third premise seems obvious to those who reflect on the kind of duplication argument proposed above. Any experiential evidence we have for accepting that we are not deceived by an evil demon would be exactly duplicated by the efforts of the evil demon and so, if we were deceived, we would have the same experiential evidence for thinking that we were not deceived as we now do. How can experiential evidence that fails to distinguish the case in which we are deceived from the case in which we are not deceived suffice for knowing that we are not deceived? Clearly it cannot, according to many philosophers. We will eventually argue against this view and defend our knowledge of such matters, but the plausibility of this line of reflection must be acknowledged. Let us consider where it leads.

One who has arrived at this stage of reflection and is prepared to affirm the first premise and the third premise can only escape from the trilemma by denying the second premise, namely, if I know that I see a hand in front of me, then I know I am not deceived by the evil demon. The formulation of the second premise in the first person is important, for the truth of it depends on my reasoning in the following way about what I know. I know that I see a hand before me. I know, as well, that if I see a hand before me, then I am not deceived by an evil demon. (The evil demon does not permit us to see any material objects but instead deceives us by producing the experiences—the sensations—we would experience if we were seeing a table in front of us when we do not. So it follows logically that if I am seeing a table in front of me, then I am not deceived by the demon.) Thus, if I know that I see a hand before me, then I know that I am not deceived by the demon.

The reasoning presented above is intended to be the reasoning of philosophical reflection. The trilemma arises for a person who has reflected on the consequences of the evil demon hypothesis and what follows from it concerning perception of the external world. Thus, the second premise is the result of someone concluding, as I have, and perhaps you have as well as you followed my reasoning, that if I know that I perceive a hand, then I know that I am not deceived by the evil demon. The trilemma arises out of philosophical reflection.

The solution of some externalists, Dretske and Nozick, for example, is to deny the second premise. According to Dretske, as we have noted, a person only needs to rule out relevant alternatives, those that are sufficiently probable to merit our consideration. It is not a relevant alternative to my seeing a hand that I am now asleep and dreaming or that I am deceived by a demon. So consider the hypothetical claim in premise 2:

HKK. If I know that I see a hand in front of me, then I know that I am not deceived by a demon.

That might seem to follow from my knowing the simpler hypothetical claim

H. If I see a hand in front of me, then I am not deceived by a demon.

According to Dretske, however, that is a mistake. The reason is that the relevant alternatives to the knowledge claim that I see a table in front of me are different from the relevant alternatives to the claim that I am not deceived by a demon. So, even though I know H, it does not follow from this that HKK is true.

The crux is that I cannot rule out that I am deceived by a demon, which is relevant to my knowing that I am not deceived by a demon but not relevant to my knowing that I see a hand in front of me. Consequently, HHK is false because I know that I see a hand in front of me, as the first premise of the trilemma asserts, even though I do not know that I am not deceived by a demon, as the third premise asserts. The alleged solution is to deny the second premise of the trilemma. To deny this premise requires denying the closure principle, however, for it is conceded that I know at one and the same time that I see a hand in front of me and that if I see a hand in front of me, then I am not deceived by the demon. Given closure, I would, as a result, know that I am not deceived by a demon. Thus, one solution of the trilemma, which follows from the externalist views of Dretske and Nozick, is to reject the closure principle and the second premise of the trilemma.

Contextualism: Another Solution

Another solution of the trilemma that accepts the first and third premises is that of contextualists such as Stewart Cohen, Keith DeRose, and David Lewis.[18] Though the accounts of these philosophers differ in important ways, they share the thesis that the standards for the evaluation of a knowledge claim depend on pragmatic features of the context. In an ordinary context, no one would deny that I know that I see a hand in front of me on the grounds that I do not know that I am not deceived by a demon if for no other reason than that no one raises questions about evil demons deceiving us as we go about the ordinary affairs of life. However, in a philosophical context, which is extraordinary because in philosophy we do consider demons and other grand deceivers, one would deny that I know that I see a hand in front of me on the grounds that I do not know that I am not deceived by a demon.

How does this distinction between ordinary contexts of evaluation and extraordinary philosophical contexts of evaluation enable us to resolve the trilemma? The resolution, according to the contextualist, is to note that when one affirms the first premise—that I know that I see a hand in front of me—one is affirming this of me claiming to know in an ordinary context in which it is not relevant to consider deceptive demons, whereas when one affirms the third premise, that I do not know that I am not deceived by a demon, one is affirming this of me in an extraordinary context in which it is relevant to consider deceptive demons.

On such accounts, the use of the word 'know' is like that of indexical expressions like 'here.' Without any change in the meaning of the word 'here' we use it to refer to different places, and it is indexed to those places. Similarly, according to the contextualist, the use of 'know' is indexed to a context in which it is used. As long as the context is held constant in our use of the word, we may affirm the second premise, though if we shift the way we think about the context as we reflect first on the antecedent—*if* I know I see a table in front of me—to the consequent—*then* I know I am not deceived by the demon—we might affirm the antecedent and deny the consequent. However, the principal claim of the contextualist is that it is only as a result of shifting the context of evaluation of the knowledge claim that the trilemma arises. If the context is held constant in our evaluation of the knowledge claim, we will either affirm the first premise and deny the third or deny the first and affirm the third, admitting the second premise. When we eliminate contextual inconstancy, the trilemma does not arise.

It is important to notice that there is an important difference between Dretske and the contextualists. On Dretske's account, there is no mistake, no contextual inconstancy, no fallacy in affirming both the first and third premise of the trilemma, that is, that I know that I see a table in front of me

and that I do not know that I am not deceived by the demon. Both those premises are true, and the culprit, the false premise, is the second one. None of this depends on context. It depends on the relevance of alternatives rather than on contextual constraints on what we should affirm about what people know.

Knowing Skeptical Hypotheses Are False

Nevertheless, both accounts agree that the third premise is true when we consider the matter as the philosophers that we are. For us, at least in our philosophical reflections, they affirm that we do not know that we are not invincibly deceived by an evil demon and, for that matter, that we do not know that we are not hallucinating or dreaming in a way that invincibly deceives us. Should we agree with them and their theories that we do not know that we are not deceived in these ways? On the contrary. I know that I am not now dreaming. I know that I am not now hallucinating. I know that no Cartesian demon deceives me. I know that no powerful scientist has my brain in a vat in his laboratory. I know that there is no braino inserted in my brain.

I may find it difficult to prove to a skeptic that I know these things. I am, however, personally justified in accepting that the skeptical hypotheses are false. My acceptance of their falsity is based on what I accept, including my acceptance that I am trustworthy in a successfully truth-connected way. If, moreover, this personal justification is undefeated and my acceptance trustworthy in a truth-connected way as I suppose, then I know that these skeptical hypotheses are all false. The skeptical hypotheses are relevant, contrary to Dretske. They are genuine objections, but they are answered by my evaluation system, and the answer is sustained in my ultrasystem to yield knowledge. This knowledge does not result from the irrelevance of the skeptical alternatives but from my being personally justified in accepting that I am not dreaming, hallucinating, deceived by an evil demon, or manipulated by a scientist. You are in the same situation, moreover. Finally, because we are right in accepting these things, our justification is irrefutable, undefeated, and converts to knowledge.

The skeptic provides an objection to our various claims to knowledge—to the claims of perception, memory, and introspection. She shows that it is possible that we are in error, and she is right in this. It is possible. We may go further and admit not only the logical possibility that we may err but also that we are genuinely fallible in what we accept. We make genuine errors of perception, memory, and introspection. Consequently, there is always some chance of error in what we accept, however small and not worth worrying about in our daily transactions, as the contextualist insists. Although we thank the skeptic for reminding us that a sound epistemology must acknowledge that we sometimes err and are ever fallible in our judg-

ment, we may at the same time neutralize her objection. Let us remind ourselves how.

Neutralizing the Fallibility Objection of the Skeptic

We acknowledge that we are fallible and there is some chance of error, however negligible, in perception, memory, and introspection, but we also accept that we are trustworthy in a truth-connected way in what we accept. The objection based on our fallibility is neutralized by our trustworthiness. It is as reasonable to accept both that we are fallible and that we are trustworthy in a truth-connected way as it is to accept only that we are fallible. It is then as reasonable for us to add that we are trustworthy in this way to the objection and accept both the possibility of error and our trustworthiness in avoiding error in a truth-connected way as to accept the skeptical worry alone.

It is, therefore, our trustworthiness that neutralizes the skeptical worries. In those instances in which we are trustworthy in a truth-connected way our justification may be undefeated by local errors and convert to knowledge. In those instances in which we accept that we are trustworthy when we are not, on the other hand, the neutralization fails in the ultrasystem and our justification is defeated. The possibility or even some small risk of error does not bring the skeptic victory, however. The small risk of error may be worthwhile in the quest for truth. One can be both fallible and trustworthy. You do not have to be perfect and infallible to be virtuous and trustworthy.

Why not Closure?

Why not, however, reject the closure principle and refute the skeptic twice over? Her ability to survive criticism has given her greater longevity than Methuselah, after all, and a double refutation seems appropriate. The problem is that rejection of the closure principle yields problematic results concerning other matters. Of course, as Gilbert Harman has noted, the mere deduction of some result from what one knows, even if one knows the deduction is valid, does not ensure that one will know the thing deduced.[19] What one knows at one time, one may fail to know at a later time because of what transpires in the interval. Deduction itself takes time. It is a process. As a result, one might in the process of deducing consequences from premises one accepts decide it was a mistake to have accepted the premises, given what follows from them, and reject the premises rather than accept the conclusion. However, this is, in fact, irrelevant to one form of the closure principle.

The closure principle may be formulated to describe what one knows at a given point in time rather than as a principle about a process that occurs over time. As a synchronic principle of knowledge, that is, as a principle

about what one knows at a *single time*, it seems correct. The principle says that if one knows at one and the same time that p and that if p then q all together, then one knows that q at that time as well.

The problem that arises from denying this principle is illustrated by an example from Saul Kripke[20] based on an earlier example from Carl Ginet.[21] Suppose at one time, I know that I see a blue barn and that if I see a blue barn, then I know that I see a barn at the same time. How could I know the former and not the latter? If we deny the epistemic closure principle, then I might. Moreover, if externalist theories were correct, it also might be the case that I might. We may illustrate the connection by supposing that I am driving in a part of the country where a clever stage builder put up barn facades here and there, which to the unsuspecting look exactly like barns. Suppose, however, that no such facades are blue, and that I, innocent of the industry of facade builders, see a blue barn. Imagine, moreover, that there are no other real barns in the area, only numerous red barn facades, and that I would not be able to tell the difference between such facades and a barn as I drive along.

Do I know that I see a blue barn? It would seem that I do not, since I cannot here tell a barn from a barn facade in my present circumstances. Notice, however, that I would not believe that I see a blue barn if I did not see a blue barn, for there are no blue barn facades. My belief tracks truth, as Nozick requires. Were tracking truth sufficient for knowledge, I would know that I see a blue barn. Notice, however, that if I also believe I see a barn, this belief would not track truth. Since there are many barn facades, it would be incorrect to say that I would not believe that I see a barn if I did not see a barn. I might believe I see a barn because I see a barn facade. Thus, if tracking truth were sufficient for knowledge, I would know that I see a blue barn but not know that I see a barn. Closure would fail.

The foregoing problem might perhaps be avoided by some modification of externalism, but it is naturally avoided by the account of knowledge as undefeated justification that we have offered. Suppose that I see a blue barn, ignorant of the existence of barn facades, as in the example. Then I will accept that I can tell whether or not I am seeing a barn in the present circumstances. This acceptance is false, however. When this acceptance is eliminated in my ultrasystem, my justification for accepting that I see a blue barn, as well as for accepting simply that I see a barn, will fail to convert into irrefutable justification. The reason is that the objection that I cannot tell whether or not I am seeing a barn in the present circumstances cannot be met, answered or neutralized, on the basis of the ultrasystem. Put in terms of the ultracritic in the ultra justification game, the objection of the ultracritic to the effect that I cannot tell whether or not I am seeing a barn in the present circumstances cannot be met by the claimant. The ultracritic wins.

The result of our theory is that I do not know that I see a blue barn anymore than I know that I see a barn in a barn facade–infested environment. The moral is that if we try to escape from skepticism by rejecting the closure principle, we may find ourselves committed to saying that we know that we see a blue barn when we do not know that we see a barn. For this reason, when the externalist replies to the skeptic that we know that we see a barn when we do not know that we are not dreaming we see a barn, he can hardly expect any more tolerant response from her than a smile of unknowing contempt.

Contextualism and Trustworthiness

Though we have rejected the form of contextualism that would yield the result that I know that I see a hand in front of me but do not know that I am not deceived in accepting this, there is some point to contextualism, as Williams has argued.[22] The point is that whether intellectual virtue and trustworthiness succeed in being connected with truth will depend on the circumstances. If we are globally deceived, then no matter how trustworthy we might be, in these circumstances, in this context, our justification for accepting what we do will be globally refuted and defeated. If we are locally deceived, then, again, no matter how trustworthy we might be in accepting what we do, the truth connection will fail, and our justification will be locally refuted and defeated. In this way, then, we find some point to contextualism, though not the kind that permits us to say that whether we know depends on the sort of pragmatic features of the context to which some contextualists, DeRose and Lewis, for example, appeal. The point is that the circumstances determine whether our trustworthiness is successfully truth connected.

Is the general reliability of trustworthiness sufficient for the kind of successful truth connection we require to convert personal justification to knowledge? No, it is necessary but not sufficient. The reason is that our trustworthiness may be generally reliable in leading us to accept what is true and avoid accepting what is false but may fail to explain why what we accept is true in the particular instance. Consider a person, Mr. Goodsumer, who is trustworthy in a reliable way in the manner in which he adds up numbers, the way he carefully double-checks what he does, but in a particular instance, though he proceeds in the usual trustworthy and reliable way, he makes a mistake, two perhaps, and gets the right answer by luck.

Mr. Goodsumer accepts that his trustworthy way of accepting the answer is successfully truth connected, but in this case, it is not, even though what he accepts as the sum is correct. His trustworthy way of doing sums is generally reliable, and he almost never makes mistakes. But in this instance, he

made mistakes and by luck got the right answer anyway. The problem is that he did not get the correct answer because he added the numbers in a trustworthy way. He was not successful because he proceeded in a trustworthy way that is reliable. Mr. Goodsumer was successful because of luck.

The explanation of why he does not know in our account of knowledge is easy enough to provide. He accepts that he is successful in getting the right answer because he has proceeded in a trustworthy and reliable way. But this is false—that is not why he was successful in getting the right answer—and the falsity of it suffices to refute and defeat his justification. The only problem is to explain what it means to say that a person is successful in accepting what is true because she has proceeded in a trustworthy and reliable manner. We have explained what it is that makes a person trustworthy in what she accepts. It is to have a disposition of a certain sort aiming at accepting what is true and avoiding accepting what is false. What makes a person who is trustworthy in what she accepts successfully reliable as well? Being successfully reliable is to succeed generally in achieving the goal of accepting what is true and avoid accepting what is false. It is, therefore, to have a high frequency of success in accepting what is true and avoiding accepting what is false, perhaps resulting from an underlying propensity toward that frequency.

What does it mean to say that trustworthiness in what one accepts is successfully connected with truth in what one accepts in a particular case? It cannot mean, as we have noted in the case of Goodsumer, that being trustworthy in what one accepts is generally or reliably successful. It means that the person is successful in accepting what is true *because* she accepts what she does in a trustworthy way in the particular case. Her trustworthiness explains her success in accepting what is true. The explanation of her success in that particular case is her trustworthiness. It is the manifestation of her disposition to be trustworthy and the underlying propensity of her disposition to lead her to accept what is true and avoid accepting what is false that explains why she accepted something true in this instance. Her trustworthiness and the reliability of it explains her success in the particular case.

The Insufficiency of Reliability and the Role of Explanation

It may be important to notice that reliablism alone will not suffice for the sort of justification required for knowledge, first, as already noted, because reliability may be opaque to the subject. The trustworthiness of the subject that involves the positive evaluation of a target acceptance on the basis of the background system, the evaluation system, provides the transparency

needed for justification. Another equally important reason why reliablism will not suffice is that global reliability might be irrelevant locally. Consider again the case of Mr. Goodsumer, who sums reliably but not in the particular case. He is not trustworthy in the way he sums in this case. Trustworthiness must be connected with truth in order for personal justification to convert to knowledge in the particular instance.

What is the required connection? It is explanation. This is the remaining insight of the explanatory coherence theory discussed in Chapter 5. Being trustworthy in the particular instance *explains* why one succeeds in obtaining truth. The way of trustworthiness is connected with the way of truth in this particular instance. I must succeed in accepting what is true in the particular case *because* I am intellectually virtuous and trustworthy in what I accept in this case.

It is not enough to point to some general success to explain why I succeed—to the general probability, however objective and lawlike, to explain why I succeed. The explanation of success in obtaining truth depends on the features of the particular case. A scientific explanation of why we succeed in obtaining truth in what we accept in a particular instance may admit of the appeal to an objective probability or scientific law, but this is not essential. We know ordinary simple things, that we see a table, that we have done a simple sum correctly, that we are speaking with another. Further, we know that we know our trustworthiness in what we accept is connected with truth in the particular instance before we can provide any general explanation of why this is so. We often know that x occurs because y does and that the occurrence of y explains the occurrence of x when we cannot provide any scientific explanation of why it is so. Science may demand such an explanation, and it is much to be desired, but epistemology can answer the skeptic without it.

Answering the Skeptic: A Concluding Discourse

What sort of answer can we give to the skeptic who affirms that we are ignorant by appeal to skeptical hypotheses about demons, scientists, dreams, or hallucinations providing invincible deception? First, we must be fair with her. Some are inclined to say that she contradicts herself when she affirms anything, for example, that no one knows anything, for in so doing she claims to know what she affirms. But this reply to the skeptic is incorrect and unfair. We often say things, tell stories, or even remark on what we think and believe without claiming to know. We are capable of epistemic modesty when we tell another what we believe. We can say, "I am not claiming to know this is true, but here is what I think about it." And the modest skeptic can proceed in this manner.

What can we say to the modest skeptic? How much we can say depends on what she allows. She herself is going to appeal to what she accepts in the discussion, for example, when she says that the invincible deception of the demon is possible. She need not claim to *know* she accepts these things, but she is going to appeal to her acceptance of them. The only other alternative is for her to say nothing, to put her hand over her mouth in silence, allowing no appeal to what she and we accept. That alternative reduces us both to silence; we can wiggle a finger or shake a head, but we say nothing.

Imagine, however, that we are confronted with a more loquacious skeptic who chooses to discuss the merits of skepticism with us. Let me put this in the first person as I engage the skeptic, and you can identify with me or the skeptic as you choose. I choose the first-person form to make it explicit that I appeal to my evaluation system, to what I accept and how I reason, to reply to the skeptic. You have your own evaluation system and may answer differently. As I reply to the skeptic, I suggest to you the form of an evaluation system that allows me to answer the skeptic. You might answer her differently, and my answer does not preclude others. If, however, you concede the correctness of what I say to justify myself, then I ask you to attend to the conclusion, namely, that you concede that my justification is irrefutable and undefeated and thus converts to knowledge. Please feel free to wear the cloak of the skeptic as I proceed and conceal your identity as an antiskeptic if that is what you are. The skeptic will forgive you.

Acceptance and Self-Trust: A Reply to the Skeptic

Here, then, is my reply to you, the skeptic. I must appeal to what I accept, for that represents my best effort, my most intellectually virtuous and trustworthy attempt to succeed in accepting what is true and avoiding accepting what is false. You might not think much of my efforts, but you will understand that I have no better way to proceed. So what do I accept? I accept many things, but let me be candid and admit that I immediately confront an either/or concerning what I accept. Either what I accept is worthy of my trust, that is, I am trustworthy at least for myself in what I accept, or I am not. If I do not accept that I am worthy of my own trust in what I accept, then I should trouble you no further; I should not trust it and should leave you in skeptical repose.

On the contrary, however, I accept that I am worthy of my trust in what I accept. I make no claim to infallibility in what I trust. Though I should prefer to be infallible, I agree with you that we are all fallible in what we accept, even what we accept about the most obvious things about what we feel, about what we think, about what we perceive, about what we remem-

ber, and so forth. I claim that I am, nevertheless, worthy of my trust in what I accept about such matters. So I place my trust in what I accept. Consider this analogy. If I am seeking to find something in a complicated city, Istanbul, for example, I do well to place my trust in a trustworthy guide even though trustworthy guides are fallible too and sometimes, however rarely, lose their way.

So I place my trust in what I accept. Shall I also place further conclusions in my reasoning? I accept that I am worthy of my trust in how I reason, that is, I am trustworthy in how I reason. I note in my defense that my acceptance and reasoning represent my best efforts to achieve my goals of accepting and concluding what is true and avoiding accepting and concluding what is false.

My acceptance and reasoning represent my attempts to proceed in an intellectually virtuous way to achieve my purpose of obtaining truth and avoiding error. Acceptance is something different from casual belief, as reasoning is something different from casual inference. Acceptance and reasoning manifest these intellectual virtues and that is why I accept that I am worthy of my own trust—trustworthy—in what I accept and how I reason.

Thus, I appeal to what I accept for the purposes of justification. If you now ask me whether I am deceived in the ways that skeptics imagine, I must tell you that I accept that I am not so deceived. I accept that I see a table in front of me. You may raise some objections, that I am dreaming, for example. But I must tell you that I accept that I am awake now, not asleep and dreaming. There is nothing in my experience that suggests a dream, for there is a continuity and coherence in my experience that is typically lacking in dreams. You say I am fallible and could be deceived in this. I admit that I am fallible and could be deceived, but I accept that I am not. You might suggest, instead, that I am simply undergoing some very systematic hallucination. But I accept that I am not hallucinating. I have taken no odd substances, I am not in any abnormal state, a lack of sleep, for example, that would produce such hallucinations. Or you might, moving to global deception, suggest that I might be deceived by the evil demon. Of course, a sufficiently powerful demon could deceive me, but I accept that I am not so deceived even if I could be. We could go on in this way, but it is better to avoid tediousness and repetition. It is what I accept that answers your objections, all of them, to personally justify me in some of the things I accept.

Moreover, though you could be right and I could be mistaken in what I accept, for I agree I am fallible, suppose for a moment that I am not mistaken. I have justified what I accept by meeting your skeptical objections as well as objections of my own. Now, to draw the conclusion that draws me, suppose I am right in what I accept, that is, what I have accepted to meet these objections is true. Then my justification will be irrefutable, for I made

no mistake anyone could use to refute me, and irrefutable justification is knowledge. That is why I conclude that I know.

No Proof That the Skeptic Is Wrong

Now you will, of course, notice that in all this I have appealed to the premise that I am worthy of my trust in what I accept. That is, of course, something I accept, and I have my reasons on the virtuous way I accept what I do. If you doubt that I am trustworthy in what I accept, I cannot prove to you that I am, and I must concede to you that I have not proven to you that you are wrong to deny that we have knowledge. Your agnoiology, your theory of ignorance, is not something I can prove to you to be false without begging the question against you. However, I would like you to consider why I accept that I know that you are wrong nonetheless.

Knowing the Skeptic Is Wrong: An Explanatory Loop

My argument loops back to the premise of my trustworthiness to explain my reasonableness in accepting what I do. The explanatory power of the argument depends on the truth of the premise of my trustworthiness. Suppose, as I accept, that I am trustworthy in what I accept and in how I reason. I can then explain my reasonableness in accepting what I do, including my trustworthiness in what I accept and how I reason. Moreover, it is perfectly consistent with this explanation to admit that the particular instances of things I accept in a trustworthy way may confirm the major premise of my trustworthiness. It is familiar feature of explanation according to scientific method that a hypothesis explaining something, for example, a principle of gravitational force explaining the motion of the planets, is confirmed by particular instances of what it explains, for example, by the observed motion of the planets as predicted by the principle. So when I explain the reasonableness of what I accept by appeal to my general trustworthiness, I may by the same scientific method go on to confirm my general trustworthiness by the trustworthiness I exhibit in particular instances of acceptance.

Two questions remain, however. The first question is whether my trustworthiness is successfully connected with truth. If it is not, that would refute my justification for what I accept and block the conversion of justification to knowledge. The second question is whether a theory of justification should permit a loop, in short, whether the loop is an intellectual virtue or vice. Some skeptic might argue that this kind of loop is a defect in a theory of reasonableness, justification, and knowledge, and the presence of it is the basis for an agnoiology of universal ignorance.

Why My Trustworthiness Is Truth Connected

Is my trustworthiness successfully connected with truth? I accept that it is and for the good reason that I have found it to be so connected with great reliability. I am trustworthy in what I accept. So, I am trustworthy in accepting that my trustworthiness is truth connected. The hypothesis of my trustworthiness in what I accept explains why I am reasonable to accept what I do, by an earlier argument, and here we need only extend the argument by noting that the *truth* of the hypothesis that my trustworthiness is successfully truth connected explains why my trustworthiness in accepting it leads me to the truth. Trustworthy acceptance leads me to the truth because my trustworthiness is successfully truth connected. My success in obtaining truth is explained by the very truth connectedness of my trustworthy acceptance.

Of course, the premise of truth connectedness must be true for the argument to be sound. What is crucial is that the truth of the premise explains both the reasonableness of accepting it and the truth of accepting it in a way that provides for irrefutable justification and knowledge of the truth of it. The loop of trustworthiness and reasonableness widens to explain the reasonableness and truth of what I accept, including that my trustworthiness in what I accept is truth connected.

If it be conceded that the widening loop is explanatory of the reasonableness of my accepting my trustworthiness, the successful truth connectedness of my trustworthiness and, indeed, the truth of these matters, it may yet be doubted by you, the skeptic, that this is a legitimate way to proceed in constructing a theory of justification. I admit that the loop precludes demonstration or proof that a skeptic is wrong. But does it permit a legitimate explanation of the reasonableness, justification, and truth of what we accept about our trustworthiness and the truth connection? Is the explanatory loop virtuous or vicious in a theory of justification?

The Virtuous Loop Maximizes Explanation

Here is my argument for concluding that the loop is virtuous. When we construct a complete theory of justification, a special issue arises when we ask whether the theory itself is justified. There are two possibilities. One is that the theory of justification explains why we are justified in accepting the theory of justification itself. The other is the theory does not explain why we are justified in accepting the theory. If our theory of justification does not explain why we are justified in accepting it, what explanation can we give? We must choose one of three alternatives. First, we might proceed to a new theory of justification to justify accepting the old one and then, alas,

face the problem anew for the new theory, falling into a regress to which there is no end. Or, second, to avoid the circle without regress we can just insist that the theory of justification is justified, though we have no rational explanation of why. This option would leave us with an unexplained surd. We can, to avoid the loop, fall into a regress of explanation without end, or affirm that the theory is an unexplained surd of explanation. The third alternative is to affirm the loop and accept that a theory of justification should explain why we are justified in accepting it.

A person seeking to maximize explanation—to explain all that one can and leave as little unexplained as one must—will prefer that a theory of justification loop back into itself and explain why we are justified in accepting it rather than leave the matter unexplained. The quest for explanation is neither necessary nor sufficient for justification, as already noted.[23] However, when the quest to maximize explanation is combined with a quest to obtain truth and avoid error, the result is admirably useful. The quest for explanation leads to the formulation of useful and powerful theories and hypotheses. The quest to obtain truth and avoid error leads to the testing and correction of errors in the explanatory theories and hypotheses. That is why I seek to maximize explanation, because of the usefulness. The loop in the theory of justification is a virtuous one for the purpose of explaining as much as we can and leaving as little unexplained as we must.

I conclude, therefore, with the observation that I have no proof that the skeptic is wrong. However, I am justified in accepting many of the things I accept, including my trustworthiness and the truth connectedness of it. Moreover, my justification is irrefutable because of the truth of what I accept. My undefeated justification converts to knowledge. I know that I am trustworthy in what I accept and that my trustworthiness is truth connected. Finally, I know that I know this, for my justification for accepting that I know is also irrefutable. The justification is undefeated and converts to knowledge. I know that I know. It is embarrassing to know, indeed, to know that I know this as well as more specific things about what I feel, perceive, remember, and infer, and not to be able to demonstrate any of it to a skeptic. Yet I am content to know and know that I know that a skeptic is wrong, even if I cannot prove it. It is, after all, knowledge and not the proof of it that I seek.

Summary

In summary, we may in our quest for truth become confident of some modest success and communicate our confidence to others by affirming that we know. We may then proceed to justify that claim to other inquirers. We thus elicit their rejoinders and sometimes change what we accept as a re-

sult. By so doing, we hope to correct what we accept and come to know our world. It is the purpose of our theory of knowledge and justification to explicate the product of this uncertain epistemic adventure. One necessary step in this explication has been to repudiate the dogmatic prejudice that we often proceed without any chance of error. Our epistemology approaches the agnoiology of skepticism without embracing the skeptical conclusion. We affirm that there is no security against failure or logical guarantee of success in our search for truth. The nobility of our objective must suffice to sustain our quest. We are, nevertheless, correct in enough of what we accept about ourselves, the external world, our trustworthiness, and the truth connectedness of it. So we may, contrary to the skeptic, know what we think we do, including the falsity of her ingenious hypotheses. We should, however, have the modesty to concede that we do not know for certain that we are right, nor can we demonstrate that she is in error. She is the touchstone of sound epistemology and merits our conscientious regard even as we affirm our knowledge that, contrary to her allegations, we know and know that we know much, though not all, of what we claim to know. The skeptic, though she may be in error, reminds us of our fallibility and of the possibility of being deceived. Paradoxically, she helps us avoid error and obtain the knowledge she denies that we have. In epistemology, as in other aspects of human endeavor, those who challenge us enable us to obtain our objectives. The skeptic is our most useful critic who protects the justification of what we accept from refutation and defeat and transforms it into the irrefutable and undefeated justification we seek. That is what the knowledge game is like.

Introduction to the Literature

There are many important works on skepticism. The most influential traditional work was probably *Meditations,* by René Descartes, and the most important traditional defense of common sense against skepticism was *Essays on the Intellectual Powers of Man,* by Thomas Reid. The most important twentieth-century article on skepticism is G. E. Moore's "A Defense of Common Sense." Some important recent books concerned with skepticism include *The Significance of Philosophical Scepticism,* by Barry Stroud; *Ignorance: A Case for Scepticism,* by Peter Unger; *Skepticism,* by Nicholas Rescher; *Certainty: A Refutation of Scepticism,* by Peter Klein; and John Greco, *Putting Skeptics in Their Place: The Nature of Skeptical Arguments and Their Role in Philosophical Inquiry.* There are important articles on skepticism in the volume *Essays on Knowledge and Justification,* edited by George S. Pappas and Marshall Swain; *Doubt,* edited by Michael Roth and Glenn Ross; and *Skepticism: A Contemporary Reader*, edited by Keith DeRose and Ted Warfield.

Notes

1. Ernest Sosa, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991); Alvin Goldman, *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge: MIT Press, 1992); Linda Zagzebski, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge: Cambridge University Press, 1996); and John Greco, *Putting Skeptics in Their Place: The Nature of Skeptical Arguments and Their Role in Philosophical Inquiry* (Cambridge: Cambridge University Press, 2000).

2. Michael Williams, *Unnatural Doubts* (Princeton: Princeton University Press, 1996).

3. Keith DeRose, "Contextualism and Knowledge Attributions," *Philosophical and Phenomenological Research* 52 (1992): 913–29; and "Solving the Skeptical Puzzle," *Philosophical Review* 104 (1995): 1–52.

4. The author defended the deepest skepticism in "Why Not Scepticism?" in *Philosophical Forum* 2 (1971): 283–98; and Peter Unger defended a very deep, if not the deepest, form in "A Defense of Skepticism," *Philosophical Review* 80 (1971): 198–219. See also Unger, *Ignorance: A Case for Skepticism* (Oxford: Clarendon, 1975); Barry Stroud, *The Significance of Philosophical Scepticism* (Minneapolis: University of Minnesota Press, 1981); Nicholas Rescher, *Skepticism: A Critical Reappraisal* (Totowa, N.J.: Rowman & Littlefield, 1980); and Peter Klein, *Certainty: A Refutation of Scepticism* (Oxford: Clarendon, 1984).

5. Thomas Reid, *The Philosophical Works of Thomas Reid,* ed. Sir William Hamilton, 8th ed. (Edinburgh: James Thin, 1895), 234; and G. E. Moore, "A Defense of Common Sense," in *Contemporary British Philosophy,* ed. J. H. Muirhead, 2d series (London: Macmillan, 1925), 193–223.

6. The most famous, of course, is René Descartes in *Meditations* I.

7. Paul Churchland, *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain* (Cambridge: MIT Press, 1995).

8. René Descartes *Meditations* I.

9. Hilary Putnum, *Reason, Truth, and History* (Cambridge: Cambridge University Press, 1984).

10. James W. Cornman, Keith Lehrer, and George S. Pappas, eds., *Philosophical Problems and Arguments: An Introduction,* 3d ed. (New York: Hackett, 1987), chap. 2.

11. See Ernest Sosa, *Knowledge in Perspective*; Linda Zagzebski, *Virtues of the Mind*; and John Greco, *Putting Skeptics in Their Place.*

12. Stewart Cohen, "Justification and Truth," *Philosophical Studies* 46 (1984): 279–96.

13. Ernest Sosa, "Philosophical Skepticism and Externalist Epistemology," in *Proceedings of the Aristotelian Society Supplementary Volume* (1994); and Sosa, "How to Defeat Opposition to Moore," *Philosophical Perspectives,* 2000.

14. Cf. Sosa, "Philosophical Skepticism and Externalist Epistemology."

15. See Fred Dretske, "Epistemic Operators," *Journal of Philosophy* 67, no. 24 (1970): 1007–23; and Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981).

16. See Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981), 123–24.

17. Cf. Ernest Sosa, "How to Defeat Opposition to Moore."

18. See Keith DeRose, "Contextualism and Knowledge Attributions," *Philosophical and Phenomenological Research* 52 (1992): 913–29; Stewart Cohen, "How to Be a Fallibilist," *Philosophical Perspectives* 2 (1988): 91–123; and David Lewis, "Elusive Knowledge," *Australasian Journal of Philosophy* 74, no. 4 (1996): 549–67.

19. Gilbert Harman, "Induction," in *Induction, Acceptance, and Rational Belief*, ed. Marshall Swain (Dordrecht: Reidel, 1970), 91.

20. Saul Kripke, unpublished review of chapter 3 of Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981).

21. Though often attributed to Goldman, the example was first used by Ginet. For the attribution, see G. C. Stine, "Skepticism and Relevant Alternatives," *Philosophical Studies* 29 (1976): 254.

22. Williams, *Unnatural Doubts*.

23. In Chapter 5.

BIBLIOGRAPHY

This list is not a complete bibliography of the subject. It contains all of the books mentioned in the Introduction to the Literature sections that follow each chapter as well as other references that may be useful to students.

Ackermann, Robert. *Belief and Knowledge.* Garden City, N.Y.: Anchor, 1972.

Adler, Jonathan E. "Knowing, Betting, and Cohering." *Philosophical Topics* 14 (1986).

Alston, William. "Concepts of Epistemic Justification." *Monist* 68 (1985). Reprinted in *Empirical Knowledge: Readings in Contemporary Epistemology,* edited by Paul K. Moser. Totowa, N.J.: Rowman & Littlefield, 1986.

_____. "The Deontological Conception of Epistemic Justification." In *Philosophical Perspectives.* Vol. 2, *Epistemology,* edited by J. Tomberlin. Atascadero, Calif.: Ridgeview, 1988.

_____. "Internalism and Externalism in Epistemology." *Philosophical Topics* 14 (1986).

_____. "Level-Confusions in Epistemology." In *Midwest Studies in Philosophy.* Vol. 5, *Studies in Epistemology,* edited by P. A. French, T. E. Uehling, and H. K. Wettstein. Minneapolis: University of Minnesota Press, 1980.

_____. "Self-Warrant: A Neglected Form of Privileged Access." *American Philosophical Quarterly* 13 (1976).

_____. "Two Types of Foundationalism." *Journal of Philosophy* 73 (1976):165–85.

_____. "Varieties of Privileged Access." *American Philosophical Quarterly* 8 (1971).

_____. "What's Wrong with Immediate Knowledge?" *Synthese* 55 (1983).

Annis, David B. "A Contextualist Theory of Epistemic Justification." *American Philosophical Quarterly* 15 (1978).

Aristotle. *Metaphysics,* Z.

Armstrong, D. M. *Belief, Truth, and Knowledge.* Cambridge: Cambridge University Press, 1973.

_____. "Does Knowledge Entail Belief?" *Proceedings of the Aristotelian Society* 70 (1969–70):21–36.

_____. "Is Introspective Knowledge Incorrigible?" *Philosophical Review* 72 (1963).

_____. *A Materialistic Theory of the Mind.* London: Routledge & Kegan Paul, 1968.

Audi, Robert. *Belief, Justification, and Knowledge.* Belmont, Calif.: Wadsworth, 1988.

_____. "Belief, Reason, and Inference." *Philosophical Topics* 14 (1986).

_____. "Believing and Affirming." *Mind* 92 (1982).

Aune, Bruce. *Knowledge, Mind, and Nature.* New York: Random House, 1967.

Austin, J. L. "Other Minds." In *Logic and Language.* 2d series, edited by A. G. N. Flew. Oxford: Blackwell, 1959.

Ayer, A. J. *The Foundations of Empirical Knowledge.* New York: St. Martin's, 1955.

_____. *Philosophical Essays.* New York: St. Martin's, 1954.

_____. *The Problem of Knowledge.* Harmondsworth, U.K.: Penguin, 1957.

_____. "Verification and Experience" [1936]. In *Logical Positivism,* edited by A. J. Ayer. New York: Free Press, 1959.

Bender, John W. "The Ins and Outs of 'Metaknowledge.'" *Analysis* 48 (1988).

_____. "Knowledge, Justification, and Lehrer's Theory of Coherence." *Philosophical Studies* 54 (1988): 355–81.

Bender, John W., ed. *The Current State of the Coherence Theory.* Boston: Kluwer, 1989.

Berkeley, George. *A Treatise Concerning the Principles of Human Knowledge,* edited by Colin M. Turbayne. Indianapolis: Bobbs-Merrill, 1957.

Blanshard, Brand. *The Nature of Thought.* London: Allen & Unwin, 1939.

BonJour, Laurence. "Can Empirical Knowledge Have a Foundation?" *American Philosophical Quarterly* 15 (1978).

_____. "The Coherence Theory of Empirical Knowledge." In *Empirical Knowledge: Readings in Contemporary Epistemology,* edited by Paul K. Moser. Totowa, N.J.: Rowman & Littlefield, 1986.

_____. "Externalist Theories of Empirical Knowledge." In *Midwest Studies in Philosophy.* Vol. 5, *Studies in Epistemology,* edited by P. A. French, T. E. Uehling, and W. K. Wettstein. Minneapolis: University of Minnesota Press, 1980.

_____. "A Reconsideration of the Problem of Induction." *Philosophical Topics* 14 (1986).

_____. *The Structure of Empirical Knowledge.* Cambridge: Harvard University Press, 1985.

Brandt, R. B., and Ernest Nagel, eds. *Meaning and Knowledge.* New York: Harcourt, Brace & World, 1965.

Bromberger, Sylvain. "An Approach to Explanation." In *Studies in Analytical Philosophy,* edited by Ronald Butler. 2d series. Oxford: Blackwell, 1965.

_____. "Why-Questions." In *Mind and Cosmos,* edited by Robert Colodny. Pittsburgh: University of Pittsburgh Press, 1966.

Butchvarov, Panayot. *The Concept of Knowledge.* Evanston, Ill.: Northwestern University Press, 1970.

Carnap, Rudolf. *The Continuum of Inductive Methods.* Chicago: University of Chicago Press, 1952.

_____. *The Logical Foundations of Probability.* 2d ed. Chicago: University of Chicago Press, 1962.

Castañeda, Hector-Neri. "Knowledge and Epistemic Obligation." In *Philosophical Perspectives.* Vol. 2, *Epistemology,* edited by S. Tomberlin. Atascadero, Calif.: Ridgeview, 1988.

Chisholm, R. M. "The Evidence of the Senses." In *Philosophical Perspectives*. Vol. 2, *Epistemology*, edited by J. Tomberlin. Atascadero, Calif.: Ridgeview, 1988.

_____. *The Foundations of Knowing*. Minneapolis: University of Minnesota Press, 1982.

_____. "On the Nature of Empirical Evidence." In *Essays on Knowledge and Justification*, edited by G. S. Pappas and Marshall Swain. Ithaca, N.Y.: Cornell University Press, 1978.

_____. *Perceiving: A Philosophical Study*. Ithaca, N.Y.: Cornell University Press, 1957.

_____. "The Place of Epistemic Justification." *Philosophical Topics* 14 (1986).

_____. *Theory of Knowledge*. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

_____. *Theory of Knowledge*. 2d ed. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

_____. *Theory of Knowledge*. 3d ed. Englewood Cliffs, N.J.: Prentice-Hall, 1989.

Chisholm, R. M., and Robert Swartz, eds. *Empirical Knowledge*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Chomsky, Noam. *Language and Mind*. New York: Harcourt, Brace & World, 1968.

_____. *Reflections on Language*. New York: Pantheon, 1975.

_____. *Rules and Representations*. New York: Columbia University Press, 1980.

Churchland, Paul. *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge: MIT Press, 1995.

Clay, Marjorie, and Keith Lehrer, eds. *Knowledge and Skepticism*. Boulder: Westview, 1989.

Cohen, L. Jonathan. *The Implications of Induction*. London: Methuen, 1970.

_____. *The Probable and the Provable*. Oxford: Clarendon, 1977.

Cohen, Stewart. "How to be a Fallibilist." In *Philosophical Perspectives*. Vol. 2, *Epistemology*. Atascadero, Calif.: Ridgeview, 1988.

_____. "Justification and Truth." *Philosophical Studies* 46 (1984):279–96.

Cornman, James W. "Foundational Versus Nonfoundational Theories of Empirical Justification." *American Philosophical Quarterly* 14 (1977). Reprinted in *Essays on Knowledge and Justification*, edited by G. S. Pappas and Marshall Swain. Ithaca, N.Y.: Cornell University Press, 1978.

_____. *Materialism and Sensations*. New Haven: Yale University Press, 1971.

_____. "On Justifying Non-Basic Statements by Basic-Reports." In *Justification and Knowledge*, edited by G. S. Pappas. Dordrecht: Reidel, 1979.

_____. *Skepticism, Justification, and Explanation*. Dordrecht: Reidel, 1980.

Cornman, James W., Keith Lehrer, and George S. Pappas. *Philosophical Problems and Arguments: An Introduction*. 3d ed. New York: Hackett, 1987.

Crumley, Jack S. *An Introduction to Epistemology*. Mountain View, Calif.: Mayfield, 1999.

Dancy, Jonathan. *An Introduction to Contemporary Epistemology*. Oxford: Blackwell, 1985.

Danto, Arthur C. *Analytical Philosophy of Knowledge*. Cambridge: Cambridge University Press, 1968.

Davidson, Donald. "A Coherence Theory of Truth and Knowledge." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest LePore. Oxford: Blackwell, 1986.

_____. *Essays on Actions and Events.* Oxford: Clarendon, 1980.
_____. "The Individuation of Events." In *Essays in Honor of Carl G. Hempel,* edited by Nicholas Rescher. Dordrecht: Reidel, 1969.
_____. *Inquiries into Truth and Interpretation.* Oxford: Clarendon Press, 1984.
Davis, Wayne, and John W. Bender. "Technical Flaws in the Coherence Theory." *Synthese* 79 (1989): 257–78.
DeRose, Keith. "Contextualism and Knowledge Attributions." *Philosophical and Phenomenological Research* 52 (1992): 913–29.
_____. "Solving the Skeptical Puzzle." *Philosophical Review* 104 (1995): 1–52.
DeRose, Keith, and Ted Warfield. *Skepticism: A Contemporary Reader.* New York: Oxford University Press, 1999.
Descartes, René. *Discourse on Method.*
_____. *Meditations.*
Dicker, Georges. *Perceptual Knowledge.* Dordrecht: Reidel, 1980.
Dretske, Fred I. "Conclusive Reasons." *Australasian Journal of Philosophy* 49 (1971):1–22.
_____. "Epistemic Operators." *Journal of Philosophy* 67, no. 24 (1970):1007–23.
_____. *Knowledge and the Flow of Information.* Cambridge: MIT Press, 1981.
_____. *Seeing and Knowing.* London: Routledge & Kegan Paul, 1969.
Duncan-Jones, Austin. "Further Questions About 'Know' and 'Think.'" In *Philosophy and Analysis,* edited by Margaret Macdonald. Oxford: Blackwell, 1966.
Eberle, R., D. Kaplan, and R. Montague. "Hempel and Oppenheim on Explanation." *Philosophy of Science* 28 (1961):418–28.
Feldman, Richard. "Epistemic Obligations." In *Philosophical Perspectives.* Vol. 2, *Epistemology,* edited by J. Tomberlin. Atascadero, Calif.: Ridgeview, 1988.
_____. "Reliability and Justification." *Monist* 68 (1985).
Fogelin, Robert. *Evidence and Meaning.* New York: Humanities, 1967.
Foley, Richard. *A Theory of Epistemic Rationality.* Cambridge: Harvard University Press, 1987.
Fumerton, Richard A. *Metaphysical and Epistemological Problems of Perception.* Lincoln: University of Nebraska Press, 1985.
Gettier, Edmund, Jr. "Is Justified True Belief Knowledge?" *Analysis* 23 (1963):121–23.
Ginet, Carl. *Knowledge, Perception, and Memory.* Dordrecht: Reidel, 1975.
Goldman, Alan. *Empirical Knowledge.* Berkeley: University of California Press, 1988.
Goldman, Alvin I. "A Causal Theory of Knowing." *Journal of Philosophy* 64 (1967):357–72. Reprinted in *Essays on Knowledge and Justification,* edited by G. S. Pappas and Marshall Swain. Ithaca, N.Y.: Cornell University Press, 1978.
_____. "Discrimination and Perceptual Knowledge." *Journal of Philosophy* 73 (1976):771–91.
_____. *Epistemology and Cognition.* Cambridge: Harvard University Press, 1986.
_____. *Liaisons: Philosophy Meets the Cognitive and Social Sciences.* Cambridge: MIT Press, 1992.
_____. "Strong and Weak Justification." In *Philosophical Perspectives.* Vol. 2, *Epistemology,* edited by J. Tomberlin. Atascadero, Calif.: Ridgeview, 1988.

_____. "What Is Justified Belief?" In *Justification and Knowledge,* edited by G. S. Pappas. Dordrecht: Reidel, 1979.

Greco, John. "Internalism and Epistemically Responsible Belief." *Synthese* 85 (1990):245–77.

_____. *Putting Skeptics in Their Place: The Nature of Skeptical Arguments and Their Role in Philosophical Inquiry.* Cambridge: Cambridge University Press, 2000.

Haack, Susan. *Evidence and Inquiry: Towards Reconstruction in Epistemology.* Oxford: Blackwell, 1993.

Hamlyn, D. W. *The Theory of Knowledge.* London: Macmillan, 1971.

Harman, Gilbert. *Change in View.* Cambridge: MIT Press, 1986.

_____. "How Belief Is Based on Inference." *Journal of Philosophy* 61 (1964):353–59.

_____. "Induction." In *Induction, Acceptance, and Rational Belief,* edited by Marshall Swain. Dordrecht: Reidel, 1970.

_____. "The Inference to the Best Explanation." *Philosophy Review* 74 (1965).

_____. "Knowledge, Reasons, and Causes." *Journal of Philosophy* 67 (1970):841–55.

_____. *Thought.* Princeton: Princeton University Press, 1973.

Hartland-Swann, John. *An Analysis of Knowing.* London: Allen & Unwin, 1958.

Heil, John. *Perception and Cognition.* Berkeley: University of California Press, 1983.

Hempel, Carl G. *Aspects of Scientific Explanation.* New York: Free Press, 1965.

_____. "Deductive-Nomological vs. Statistical Explanation." In *Minnesota Studies in the Philosophy of Science.* Vol. 3, edited by Herbert Feigl and Grover Maxwell. Minneapolis: University of Minnesota Press, 1962.

Hill, Thomas. *Contemporary Theories of Knowledge.* New York: Macmillan, 1961.

Hilpinen, Risto. "Knowledge and Conditionals." In *Philosophical Perspectives.* Vol. 2, *Epistemology,* edited by J. Tomberlin. Atascadero, Calif.: Ridgeview, 1988.

_____. "Knowledge and Justification." *Ajatus* 33 (1971).

Hintikka, Jaako. *Knowledge and Belief.* Ithaca, N.Y.: Cornell University Press, 1962.

_____. "A Two-Dimensional Continuum of Inductive Methods." In *Aspects of Inductive Logic,* edited by J. Hintikka and P. Suppes. Amsterdam: North-Holland, 1966.

Hintikka, Jaako, and J. Pietarinen. "Semantic Information and Inductive Logic." In *Aspects of Inductive Logic,* edited by J. Hintikka and P. Suppes. Amsterdam: North-Holland, 1966.

Honderich, Ted. *A Theory of Determinism.* Oxford: Clarendon, 1988.

Horwich, Paul. *Truth.* Oxford: Blackwell, 1990.

Hume, David. *An Inquiry Concerning Human Understanding.* Indianapolis: Bobbs-Merrill, 1977.

_____. *A Treatise of Human Nature.* Oxford: Clarendon, 1888. Originally published in 1739.

James, William. "The Will to Believe." In *The Writings of William James: A Comprehensive Edition,* edited by John J. McDermott. Chicago: University of Chicago Press, 1977.

Jeffrey, Richard. *The Logic of Decision.* New York: McGraw-Hill, 1965.
Kant, Immanuel. *Prolegomena to Any Future Metaphysics.* Translated by Lewis White Beck. New York: Liberal Arts, 1950. Based on the Mahaffy-Carus translation.
Klein, Peter. *Certainty: A Refutation of Scepticism.* Oxford: Clarendon Press, 1984.
_____. "Immune Belief Systems." *Philosophical Topics* 14 (1986).
_____. "Knowledge, Causality, and Defeasibility." *Journal of Philosophy* 73 (1976).
_____. "A Proposed Definition of Propositional Knowledge." *Journal of Philosophy* 68 (1971):471–82.
_____. "Radical Interpretation and Global Skepticism." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson,* edited by Ernest LePore. Oxford: Blackwell, 1986.
_____. "Real Knowledge." *Synthese* 55 (1983).
Körner, Stephan. *Experience and Theory.* London: Routledge & Kegan Paul, 1966.
Kvanvig, Jonathan. "How to be a Reliabilist." *American Philosophical Quarterly* 23 (1986).
_____. "Swain on the Basing Relation." *Analysis* 45 (1985):153–58.
Kyburg, Henry E., Jr. *Probability and the Logic of Rational Belief.* Middletown, Conn.: Wesleyan University Press, 1961.
_____. *Epistemology and Inference.* Minneapolis: University of Minnesota Press, 1983.
Lehrer, Adrienne. "Semantic Cuisine." *Journal of Linguistics* 5 (1969):39–55.
_____. *Semantic Fields and Lexical Structure.* Amsterdam: North-Holland, 1974.
Lehrer, Keith. "Belief, Acceptance, and Cognition." In *On Believing,* edited by Herman Parret. Berlin: Walter de Gruyter, 1983.
_____. "Coherence, Justification, and Chisholm." In *Philosophical Perspectives.* Vol. 2, *Epistemology,* edited by J. Tomberlin. Atascadero, Calif.: Ridgeview, 1988.
_____. "The Coherence Theory of Knowledge." *Philosophical Topics* 14 (1986).
_____. "The Gettier Problem and the Analysis of Knowledge." In *Justification and Knowledge,* edited by G. S. Pappas. Dordrecht: Reidel, 1979.
_____. *Knowledge.* Oxford: Clarendon, 1974.
_____. "The Knowledge Cycle." *Nous* 11 (1977).
_____. "Knowledge, Truth, and Evidence." *Analysis* 25 (1965):168–75.
_____. "Knowledge, Truth, and Ontology." In *Language and Ontology: Proceedings of the Sixth International Wittgenstein Symposium,* edited by Werner Leinfellner. Vienna: Holder-Pichler-Tempsky, 1982.
_____. "Metaknowledge: Undefeated Justification." *Synthese* 74 (1988).
_____. "Metamind: Belief, Consciousness, and Intentionality." In *Belief: Forum, Content, and Function,* edited by R. J. Bogdan. Oxford: Clarendon Press, 1986.
_____. "Personal and Social Knowledge." *Synthese* 73 (1987).
_____. "Self-Profile." In *Keith Lehrer,* edited by R. J. Bogdan. Dordrecht: Reidel, 1981.
_____. *Self-Trust: A Study of Reason, Knowledge, and Autonomy.* Oxford: Clarendon, (1997).
_____. "Why Not Scepticism?" *Philosophical Forum* 2 (1971):283–98.

Lehrer, Keith, and Stewart Cohen. "Justification, Truth, and Coherence." *Synthese* 55 (1983):191–207.

Lehrer, Keith, and Thomas Paxson Jr. "Knowledge: Undefeated Justified True Belief." *Journal of Philosophy* 66 (1969):225–37.

Lehrer, Keith, and Carl Wagner. *Rational Consensus in Science and Society.* Dordrecht: Reidel, 1981.

Lemmon, E. J. "If I Know, Do I Know That I Know?" In *Epistemology: New Essays in the Theory of Knowledge,* edited by A. Stroll. New York: Harper & Row, 1967.

Levi, Issac. *Gambling with Truth: An Essay on Induction and the Aims of Science.* New York: Knopf, 1967.

Lewis, C. I. *An Analysis of Knowledge and Valuation.* Paul Carus Lectures. 7th series. LaSalle, Ill.: Open Court, 1946.

_____. *Mind and the World Order.* New York: Dover, 1929.

Lewis, David. "Elusive Knowledge." *Australasian Journal of Philosophy* 74, no. 4 (1996):549–67.

Locke, Don. *Memory.* New York: Anchor, 1971.

Locke, John. *An Essay Concerning Human Understanding.* New York: Dutton, 1928. Originally published in 1689.

Lycan, William. *Judgement and Justification.* New York: Cambridge University Press, 1988.

Lyons, John. *Structural Semantics.* Oxford: Blackwell, 1963.

Malcolm, Norman. "Knowledge and Belief." In *Knowledge and Belief,* edited by A. P. Griffiths. Oxford: Blackwell, 1967.

_____. *Knowledge and Certainty.* Ithaca, N.Y.: Cornell University Press, 1975.

Margolis, Joseph. "The Problem of Justified Belief." *Philosophical Studies* 23 (1972):405–09.

Martin, Robert, ed. *Recent Essays on Truth and the Liar Paradox.* Oxford: Clarendon, 1984.

McGee, Vann. *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth.* New York: Hackett, 1990.

McGinn, Colin. "The Concept of Knowledge." In *Midwest Studies in Philosophy.* Vol. 9, *Causation and Causal Theories,* edited by P. A. French, T. E. Uehling, and H. K. Wettstein. Minneapolis: University of Minnesota Press, 1984.

Mill, John Stuart. *A System of Logic.* London: Longmans, 1843.

Moore, G. E. "A Defense of Common Sense." In *Contemporary British Philosophy,* edited by J. H. Muirhead. 2d series. London: Allen & Unwin, 1925.

_____. *Philosophical Papers.* London: Allen & Unwin (1959): 226–35.

Morton, Adam. *A Guide Through the Theory of Knowledge.* Encino, Calif.: Dickenson, 1977.

Moser, Paul K. "Critical Notice of *The Structure of Empirical Knowledge,* by L. Bonjour." *Philosophy and Phenomenological Research* 47 (1987).

_____. *Empirical Justification.* Dordrecht: Reidel, 1985.

_____. "Internalism and Coherentism: A Dilemma." *Analysis* 48 (1988).

_____. *Knowledge and Evidence.* New York: Cambridge University Press, 1989.

_____. "Propositional Knowledge." *Philosophical Studies* 52 (1987).

Moser, Paul K., ed. *Empirical Knowledge: Readings in Contemporary Epistemology.* Totowa, N.J.: Rowman & Littlefield, 1986.

_____. *A Priori Knowledge.* Oxford: Oxford University Press, 1987.

Moser, Paul K., and Arnold Vander Nat, eds. *Human Knowledge: Classical and Contemporary Approaches.* Atascadero, Calif.: Ridgeview, 1988.

Nakhnikian, George. "Incorrigibility." *Philosophical Quarterly* 28 (1968):207–15.

Pappas, George S., and Marshall Swain, eds. *Essays on Knowledge and Justification.* Ithaca, N.Y.: Cornell University Press, 1978.

Parsons, Katherine Pyne. "Mistaking Sensations." *Philosophical Review* 79 (1970):201–13.

Pastin, Mark. "Modest Foundationalism and Self-Warrant." *American Philosophy Quarterly* 9 (1975).

Plantinga, Alvin. *Warrant and Proper Function.* New York: Oxford University Press, 1993.

_____. *Warrant: The Current Debate.* New York: Oxford University Press, 1993.

Plato. *Theaetetus; Symposium;* and *Phaedo.*

Pollock, John L. *Knowledge and Justification.* Princeton: Princeton University Press, 1974.

_____. "A Plethora of Epistemological Theories." In *Justification and Knowledge,* edited by G. S. Pappas. Dordrecht: Reidel, 1979.

_____. "Reliability and Justified Belief." *Canadian Journal of Philosophy* 14 (1984).

Pollock, John L., and Joseph Cruz. *Contemporary Theories of Knowledge.* Lanham, Md.: Rowman & Littlefield, 1999.

Posner, Michael. *Cognition: An Introduction.* Glenview, Ill.: Scott, Foresman, 1973.

Price, H. H. *Perception.* London: Methuen, 1932.

_____. *Truth and Corrigibility.* Oxford: Clarendon, 1936.

Prichard, H. "Knowing and Believing." In *Knowledge and Belief,* edited by A. P. Griffiths. Oxford: Blackwell, 1967.

Putnam, Hilary. *Reason, Truth, and History.* Cambridge: Cambridge University Press, 1984.

Quine, W. V. O. *From a Logical Point of View.* Cambridge: Harvard University Press, 1953.

_____. *Ontological Relativity and Other Essays.* New York: Columbia University Press, 1969.

_____. *Word and Object.* Cambridge: MIT Press, 1960.

Quine, W. V. O., and Joseph Ullian. *The Web of Belief.* 2d ed. New York: Random House, 1978.

Quinton, Anthony. *The Nature of Things.* London: Routledge & Kegan Paul, 1973.

Radford, Colin. "Knowledge—By Examples." *Analysis* 27 (1966):1–11.

_____. "Analyzing 'Knows That.'" *Philosophical Quarterly* 20 (1970):228–29.

Reichenbach, Hans. *The Theory of Probability.* Berkeley: University of California Press, 1949.

Reid, Thomas. *The Works of Thomas Reid, D.D.,* edited by Sir William Hamilton. 8th ed. Edinburgh: James Thin, 1895.

Rescher, Nicholas. *Cognitive Systematization.* Totowa, N.J.: Rowman & Littlefield, 1979.

_____. *The Coherence Theory of Truth.* Oxford: Oxford University Press, 1973.

_____. "Foundationalism, Coherentism, and the Idea of Cognitive Systematization." *Journal of Philosophy* 71 (1974).
_____. *Skepticism: A Critical Reappraisal.* Totowa, N.J.: Rowman & Littlefield, 1980.
Rosenberg, Jay. *One World and Our Knowledge of It.* Dordrecht: Reidel, 1980.
Roth, Michael, and Leon Galis, eds. *Knowing: Essays in the Theory of Knowledge.* New York: Random House, 1970.
Roth, Michael, and Glenn Ross. *Doubting.* Boston: Kluwer, 1990.
Russell, Bertrand. *An Inquiry into Meaning and Truth.* London: Oxford University Press, 1940.
_____. "Knowledge by Acquaintance and Knowledge by Description." *Proceedings of the Aristotelian Society* 11 (1910–11):108–28.
_____. *The Problems of Philosophy.* London: Oxford University Press, 1959.
Ryle, Gilbert. "Knowing How and Knowing That." *Proceedings of the Aristotelian Society* 46 (1945–46):1–16.
Savage, C. W., ed. *Perception and Cognition: Issues in the Foundations of Psychology.* Minneapolis: University of Minnesota Press, 1978.
Schlick, Moritz. "The Foundation of Knowledge" [1934]. In *Logical Positivism,* edited by A. J. Ayer. New York: Free Press, 1957.
Sellars, W. F. *Science, Perception, and Reality.* London: Routledge & Kegan Paul, 1963.
Shoemaker, Sydney. "On Knowing One's Own Mind." In *Philosophical Perspectives.* Vol. 2, *Epistemology,* edited by J. Tomberlin. Atascadero, Calif.: Ridgeway, 1988.
Shope, Robert. *The Analysis of Knowing.* Princeton: Princeton University Press, 1983.
Siegel, Harvey. "Justification, Discovery, and the Naturalizing of Epistemology." *Philosophy of Science* 47 (1980).
Slaght, Ralph. "Is Justified True Belief Knowledge? A Selected Critical Survey of Recent Work." *Philosophy Research Archives* 3 (1977).
Sosa, Ernest. "'Circular' Coherence and 'Absurd' Foundations." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson.* Oxford: Blackwell, 1986.
_____. "Epistemic Presupposition." In *Justification and Knowledge,* edited by G. S. Pappas. Dordrecht: Reidel, 1979.
_____. "Experience and Intentionality." *Philosophical Topics* 14 (1986).
_____. "How to Defeat Opposition to Moore." In *Philosophical Perspectives,* 2000.
_____. "Knowledge in Context, Skepticism in Doubt." In *Philosophical Perspectives.* Vol. 2: *Epistemology,* edited by J. Tomberlin. Atascadero, Calif.: Ridgeview, 1988.
_____. *Knowledge in Perspective: Selected Essays in Epistemology.* Cambridge: Cambridge University Press, 1991.
_____. "Nature Unmirrored, Epistemology Naturalized." *Synthese* 55 (1983).
_____. "Philosophical Skepticism and Externalist Epistemology." In *Proceedings of the Aristotelian Society Supplementary Volume* (1994).

_____. "Proper Functionalism and Virtue Epistemology." *Nous* 27, no. 1 (1993): 51–65.

_____. "The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge." In *Midwest Studies in Philosophy.* Vol. 5, *Studies in Epistemology,* edited by P. A. French, T. E. Uehling, and H. K. Wettstein. Minneapolis: University of Minnesota Press, 1980.

Stine, G. C. "Skepticism and Relevant Alternatives." *Philosophical Studies* 5, no. 29 (1976).

Stroud, Barry. *The Significance of Philosophical Scepticism.* Oxford: Clarendon, 1984.

Swain, Marshall. *Reasons and Knowledge.* Ithaca, N.Y.: Cornell University Press, 1981.

Tarski, Alfred. "The Concept of Truth in Formalized Languages." *Logic, Semantics, and Metamathematics.* Oxford: Clarendon, 1969.

_____. "The Semantic Conception of Truth." In *Semantics and the Philosophy of Language,* edited by L. Linsky. Urbana: University of Illinois Press, 1952.

Thalberg, Irving. "In Defense of Justified True Belief." *Journal of Philosophy* 66 (1969):794–803.

Tweedale, Jonathan. "Epistemic Humility and Undefeated Justification." *STOA,* 2000.

Unger, Peter. "The Cone Model of Knowledge." *Philosophical Topics* 14 (1986).

_____. "A Defense of Skepticism." *Philosophical Review* 80 (1971):198–219.

_____. *Ignorance: A Case for Scepticism.* Oxford: Clarendon, 1975.

Van Cleve, James. "Epistemic Supervenience and the Circle of Belief." *Monist* 68 (1985).

_____. "Foundationalism, Epistemic Principles, and the Cartesian Circle." *Philosophical Review* 88 (1979):55–91.

Will, Frederick L. *Induction and Justification.* Ithaca, N.Y.: Cornell University Press, 1974.

Williams, Michael. "Do We (Epistemologists) Need a Theory of Truth?" *Philosophical Topics* 14 (1986).

_____. "Inference, Justification, and the Analysis of Knowledge." *Journal of Philosophy* 75 (1978).

_____. *Unnatural Doubts.* Princeton: Princeton University Press, 1996.

Wittgenstein, Ludwig. *On Certainty.* Translated by Dennis Paul and G. E. M. Anscombe. Oxford: Blackwell, 1969.

Zagzebski, Linda. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge.* Cambridge: Cambridge University Press, 1996.

Ziff, Paul. "The Simplicity of Other Minds." In *New Readings in Philosophical Analysis,* edited by H. Feigl, W. F. Sellars, and K. Lehrer. New York: Appleton-Century-Crofts, 1972.

INDEX

Acceptance, 13, 25–26, 124–125
  and belief, 13–14, 40, 124–125
  as a condition of knowledge, 12–14
  functional role of, 39–40, 41
  and information, 37–39, 42–43
  and knowledge, 36–43
  and memory, 37–39, 42–43
  and preference, 40
  and trustworthiness. *See* Principle of trustworthiness
  and truth, 25–26. *See also* Principle of trustworthiness
  See also Acceptance system; Belief
Acceptance system, 119, 123–125
  defined, 130
  See also Evaluation system; Reasonableness; Ultrasystem
Ackerman, Robert, 233*b*
Adler, Jonathan E., 233*b*
Agnoiology, 205–206. *See also* Skepticism
Alston, William P., 69(n7), 84, 95, 96(n11), 233*b*
Analysis, philosophical, 7–11
  as meaning analysis, 7–8
  and science, 8–9
  as specification of necessary and sufficient conditions, 8–9
  explanatory role of, 10
  See also Explication
Analyticity, 51
Annis, David B., 233*b*
Aristotle, 1, 21(n2), 233*b*
Armstrong, David M., 42, 43, 44(n18), 178, 186, 203, 203(n3), 204(nn8, 16), 233*b*
Audi, Robert, 21, 174(n10), 234*b*
Aune, Bruce, 101, 121, 121(n3), 144, 149(n3), 234*b*
Austin, J. L., 33, 43, 44(n8), 234*b*
Ayer, Alfred J., 21, 22(nn11, 15, 16), 23(n20), 63, 68, 69(n6), 70(nn14, 15), 234*b*

Background system,
  and acceptance. *See* Acceptance system
  and justification, 123, 126
  See also Evaluation system
Basic belief, 15, 16, 45
  and contingent self-justification, 83–85
  fallible, 71–95
  as a guarantee of truth, 46–48, 50
  and incorrigibility. *See* Incorrigibility
  and independent information, 83, 93
  infallible. *See* Infallible belief
  and necessary truth, 79–83
  and *prima facie* justification, 46–47
  and probability, 47, 85–94
  and reliability, 47–48
  and self-justification, 48
  and semantic information, 76–79
  See also Foundationalism; Perceptual belief
Basing relation, 195–197
Belief
  and acceptance, 13–14
  as a condition of knowledge, 13
  basic. *See* Basic belief
  and causation. *See* Basing relation
  fallible. *See* Basic belief, fallible

and independent information, 72–73, 75. *See also* Basic belief
infallible. *See* Infallible belief
innocent, 73–75. *See also* Basic belief
perceptual. *See* Perceptual belief
and probability, 85–94
and reliability, 74–75
self-justified. *See* Self-justification
See also Acceptance; Observation statement
Bender, John W., 148, 234*b*, 236*b*
Berkeley, George (Bishop), 2, 22(n4), 63–64, 70(n12), 234*b*
Blanshard, Brand, 121(n2), 234*b*
BonJour, Laurence, 23(n18), 69(n4), 111, 122(nn12, 20), 144, 148, 149(n3), 234*b*
Brand, R. B., 234*b*
Bromberger, Sylvain, 106, 107–108, 112, 121(nn8, 9), 122(n15), 234*b*
Butchvarov, Panayot, 68, 69(n6), 234*b*

Carnap, Rudolf, 9, 22(n12), 96(nn13, 14), 234*b*
Castañeda, Hector-Neri, 234*b*
Chisholm, Roderick M., 19, 21, 22(nn16, 17), 23(nn20, 24), 63, 68, 69(nn1, 2), 70(n16), 75–76, 95, 95(n2), 96(nn6, 7), 111, 122(n13), 156, 174(n5), 201, 204(n24), 235*b*
Chomsky, Noam, 235*b*
Churchland, Paul, 22(n6), 231(n7), 235*b*
Clay, Marjorie, 235*b*
Cohen, L. Jonathan, 96(n12), 235*b*
Cohen, Stewart, 190, 203, 204(n19), 210, 218, 231(n12), 232(n18), 235*b*, 238*b*
Coherence
and acceptance system, 129
and consistency, 104
as explanation, 101–106. *See also* Explanatory coherence
as implication, 100–101
See also Reasonableness
Coherentism, 16–17, 97–175, 201–203
and the basic sources of knowledge, 165–166
and the circularity objection, 97–100
explanatory. *See* Explanatory coherentism
and the isolation objection. *See* Isolation objection
and the regress objection, 97–100
a schema for, 97
and self-justified beliefs, 99–100
See also Coherence; Principle of trustworthiness
Conditions of knowledge
an acceptance condition, 12–14
a fourth condition, 20. *See also* Gettier problem.
a justification condition, 14–15
a truth condition, 11–12
Contextualism, 218
and skepticism, 218–219
and trustworthiness, 222–223
Conservatism. *See* Epistemic conservatism
Cornman, James W., 204(n9), 231(n10), 235*b*
Counterexample, 9–10
Crumley, Jack S., 21, 235*b*
Cruz, Joseph, 21, 22(n17), 68, 69(nn1, 3, 7), 240*b*

Dancy, Jonathan, 235*b*
Danto, Arthur, 68, 69(n6), 235*b*
David, Marian, 44(n6)
Davidson, Donald, 60, 69(n11), 151–152, 174(n1), 235*b*
Davis, Wayne, 236*b*
DeRose, Keith, 205, 218, 222, 230, 231(n3), 232(n18), 236*b*
Descartes, René, 2, 3, 21(n3), 49, 53, 62–63, 69(n5), 140, 148(n1), 207, 210, 230, 231(nn6, 8), 236*b*
Desire, 40
Dicker, Georges, 236*b*
Dretske, Fred, 22(n10), 23(n19), 38, 43, 44(n16), 178–179, 183, 186,

189, 203, 203(n3), 204(nn8, 10, 12, 17), 214–215, 217, 218, 219, 232(nn15, 16), 236*b*
Duncan-Jones, Austin, 44(n7), 236*b*

Eberle, R., 121(n7), 236*b*
Empiricist epistemology, 16, 48–49
Epistemic conservatism, 119
Epistemology
  critical, 3–4
  and ethics, 111–112
  metaphysical, 1
  skeptical, 2
Ethics, and epistemology, 111–112
Evaluation system, 126–128
  defined, 130, 170
  and reasonableness, 129
  See also Acceptance system; Background system; Coherence; Preference system; Reasonableness; Reasoning system
Explanation, 106–108
  the deductive model of, 106–108
  an epistemic analysis of, 107–108
  as a primitive, 108
  and truth, 120–121
Explanatory Coherentism, 101–121
  and conditioned response, 110–112
  defined, 105–106
  and justification without explanation, 114–116
  the maximization of coherence and the size of the explanation basis, 108–109
  and natural selection, 111–112
  and observation statements, 109–112
  and opacity, 116
  and the problem of competing systems of belief, 104, 116–118
  and self-explanation, 112–114
  and spontaneity, 111
  and weak explanations, 116–117
  See also Epistemic conservatism; Explanation; Observation statement; Simplicity
Explication, 9
Externalism, 17–18, 177–203
  and the basing relation, 195–197
  and deductive closure, 214–217
  defined, 177
  and the invincibility objection, 190–191
  and justification, 178, 184, 189–191
  and opacity, 185–188
  and probability, 87–89
  and relevant alternatives, 181, 215
  and skepticism, 180–181, 214–217
  See also Naturalism

Fallible foundations, 46–48, 71
  compared with infallible foundations, 46
  and contingent self-justification, 83–85
  and *prima facie* justification, 81–83
  and probability, 85–94
  and semantics. *See* Semantic foundationalism
  See also Basic belief
Feldman, Richard, 201, 203, 204(n24), 236*b*
Fogelin, Robert, 121(n1), 236*b*
Foley, Richard, 69(n7), 95, 236*b*
Foundationalism, 15–16, 45–95
  fallible. *See* Fallible foundations.
  infallible. *See* Infallible foundations
  and nonbasic belief. *See* Phenomenalism
  See also Basic belief
Fumerton, Richard A., 236*b*

Galis, Leon, 174, 240*b*
Gettier, Edmund Jr., 19, 23(nn21, 23, 25), 50, 155, 156, 174, 174(nn2, 3), 236*b*. *See also* Gettier problem
Gettier problem, 18–20, 50, 155–158. *See also* Justification, undefeated
Ginet, Carl, 68, 221, 232(n21), 236*b*
Goldman, Alan, 121, 122(n14), 236*b*
Goldman, Alvin I., 23(n19), 96(n4), 175(n12), 177–178, 182, 184,

186, 188, 189, 190, 196, 203, 203(nn1, 2, 5), 204(nn10, 11, 13, 18, 20), 205, 231(n1), 232(n21), 236*b*
Greco, John, 205, 209, 230, 231(nn1, 11), 237*b*

Haack, Susan, 175(n13), 237*b*
Hamlyn, D. W., 237*b*
Harman, Gilbert, 74, 96(n5), 101, 118, 119, 121, 121(n3), 122(n19), 144, 149(n3), 159, 174(nn3, 9), 196, 204(n22), 220, 232(n19), 237*b*
Hartland-Swann, John, 22(n8), 237*b*
Heil, John, 237*b*
Hempel, Carl G., 106, 107–108, 121(n6), 122(n10), 148, 149(n4), 237*b*
Hill, Thomas, 237*b*
Hilpinen, Risto, 158, 174(n7), 237*b*
Hintikka, Jaakko, 96(n14), 149(n4), 237*b*
Honderich, Ted, 237*b*
Horwich, Paul, 43, 44(n4), 237*b*
Hume, David, 2–3, 22(n5), 178, 203(n6), 237*b*

Incorrigibility, 50–63
  defined, 51, 53
  and logical necessity. *See* Infallible belief, and logical necessity
  and meaning. *See* Infallible belief, and meaning
  and nomological necessity. *See* Infallible belief, and nomological necessity
  and opacity. *See* Infallible belief, and opacity
Infallible belief
  and inference, 58
  and irresistibility, 53
  and logical necessity, 50–51, 52
  and meaning, 59–61. *See also* Principle of charity
  and nomological necessity, 59
  and opacity, 61–62
  about sensation, 57–58
  about thought, 54–56
  and transparency, 61–62
  See also Basic belief
Infallible foundations,
  compared with fallible foundations, 46
  and justification as guarantee of truth, 49–50
  and nonbasic beliefs, 63. *See also* Phenomenalism
  See also Basic belief; Incorrigibility; Infallible belief
Information, 5–7, 10–11
  as known to be correct, 10–11, 37–39
  correct, 6, 186–188
  without knowledge, 11, 37–39
  and memory, 37–39
  possession of, 6–7, 186–188
  source of, 6
  See also Knowledge
Introspection. *See* Infallible belief, about thought.
Isolation Objection, 137
  ultra justification game and the, 163–165
Internalism, 172–173, 190

James, William, 237*b*
Jeffrey, Richard, 95, 96(n13), 237*b*
Justification, 123, 137
  and causation. *See* Basing relation
  and coherence. *See* Coherentism
  and conditioned response, 109–11
  as a condition of knowledge, 14–15
  determination of, 173
  and evidence, 98–100
  and explanation. *See* Explanatory coherentism
  and independent information, 72–73, 74–76
  and innocent belief, 73–75
  and natural selection, 111
  and nonbasic belief. *See* Phenomenalism
  and opacity. *See* Opacity
  personal. *See* Personal justification

and *prima facie* justification. *See* *Prima facie* justification
and probability, 85–94
and self-explanation. *See* Explanatory coherentism, and self-explanation
and self-justification. *See* Self-justification
and semantics, 76–77. *See also* Semantic foundationalism
the subjective approach to, 173
theories of, 15–18
and trustworthiness. *See* Trustworthiness
undefeated. *See* Undefeated justification
and use of noncomparative terms, 75–76
See also Acceptance system; Coherentism; Externalism; Foundationalism; Justification game; Principle of trustworthiness
Justification game, 132–134
and personal justification, 136–137
and reliability, 198–199
See also Ultra justification game

Kant, Immanuel, 237*b*
Kaplan, David, 121(n7), 236*b*
Keystone, 143
Klein, Peter, 158, 174(nn1, 7), 230, 231(n4), 238*b*
Knowledge
and acceptance, 32, 36–43
acquaintance sense of, 5
analysis of, 11–21. *See also* Knowledge, defined
and belief, 32–33, 34–36
borderline cases of, 36–37, 41
characteristically human, 7, 39
competence sense of, 5
conditions of. *See* Conditions of knowledge
defined, 20–21, 169–171
fallible, 208–209
information sense of, 5–7, 11
without justification. *See* Externalism
and memory, 37–39, 42–43
as a performative utterance, 33–34
See also Acceptance; Coherentism; Externalism; Foundationalism; Information
Körner, Stephan, 37, 44(15), 238*b*
Kripke, Saul, 221, 232(n20)
Kvanvig, Jonathan, 69(n3), 204(n22), 238*b*
Kyburg, Henry E., Jr., 146, 149(n5), 238*b*

Lehrer, Adrienne, 78, 96(n10), 238*b*
Lehrer, Keith, 23(nn18, 22, 25), 43, 44(n17), 121(n3), 148, 174(nn2, 6, 8), 204(nn9, 19), 231(nn4, 10), 235*b*, 238*b*, 239*b*
Lemmon, E. J., 239*b*
Levi, Isaac, 148, 149(n4), 239*b*
Lewis, C. I., 111, 122(n13), 239*b*
Lewis, David, 218, 222, 239*b*
Locke, Don, 239*b*
Locke, John, 239*b*
Logical necessity. *See* Infallible belief, and logical necessity
Lottery paradox, 146–147
Lycan, William, 101, 121, 121(n3), 122(n20), 144, 149(n3), 239*b*
Lyons, John, 78, 96(n10), 239*b*

McGee, Van, 43, 44(n5), 239*b*
McGinn, Colin, 22(n7), 239*b*
Malcolm, Norman, 239*b*
Margolis, Joseph, 174(n4), 239*b*
Martin, Robert, 43, 239*b*
Meaning
and belief, 59–61
and skepticism, 77–79
Memory
and acceptance, 37–39, 42–43
without belief, 36
and information. *See* Information, and memory
and knowledge. *See* Knowledge, and memory
Metamind, 40
Mill, John Stuart, 239*b*

Montague, Richard, 121(n7), 236*b*
Moore, G. E., 22(n15), 206, 230, 231(n5), 239*b*
Morton, Adam, 239*b*
Moser, Paul K., 21, 22(n17), 68, 239*b*

Nagel, Ernest, 234*b*
Nakhnikian, George, 69(n8)
Naturalism, 178–179
  and causation, 182–184
  and historical reliabilism, 184
  and truth tracking, 184
  See also Externalism
Naturalized epistemology. *See* Naturalism
Neutralization, 134–136
  defined, 136
Nozick, Robert 23(n19), 178, 179, 184, 186, 189, 203, 203(n4), 204(n17), 214–215, 217, 221, 232(nn15, 20)

Objection, defined, 130–131
Observation statement
  and conditioned response, 110–112
  and natural selection, 111–112
  as self-explanatory, 112–114
  and spontaneity, 111
  See also Empiricist epistemology; Explanatory coherence
Opacity, 81, 83
  and infallibility, 61–62
  and explanation, 116
  and externalism, 185–188

Pappas, George S., 21, 174, 204(n9), 230, 231(n10), 235*b*, 240*b*
Paradox
  and truth, 27–31
  lottery, 146–147
Parsons, Katherine Pyne, 69(n9), 240*b*
Pastin, Mark, 69(n7), 95, 240*b*
Paxson, Thomas Jr., 174(n8), 238*b*
Perceptual Belief, 72–76, 110
  and independent information, 72–76. *See also* Basic belief
  as innocent belief, 73–75
  and the noncomparative use of words, 75–76
Performative utterances. *See* Knowledge, as a performative utterance
Personal Justification, 127–128
  defined, 131–132, 137
  and the justification game, 132–134, 136–137
  and undefeated justification, 153, 171–173
  See also Evaluation system; Justification game
Phenomenalism, 63–67
Pietarinen, J., 149(n4), 237*b*
Plantinga, Alvin, 184, 188, 204(nn15, 21), 240*b*
Plato, 1, 21(n1), 240*b*
Pojman, Louis P., 21
Pollock, John L., 21, 22(n17), 68, 69(nn1, 3, 7), 96(n9), 240*b*
Posner, Michael, 240*b*
Poulsen, Rachael, 204(n23)
Preference, 40
  and acceptance, 40, 127
  and reason, 40
Preference system, 126–127
  defined, 130
  See also Evaluation system
Price, H. H., 240*b*
Prichard, H., 240*b*
Prima Facie Justification, 81–83
  and necessity, 81–83
  See also Basic belief; Belief, innocent; Perceptual belief; Semantic foundationalism
Principle of charity, 151–152
  and belief ascription, 60
Principle of trustworthiness, 138–144, 201–203
  See also Trustworthiness
Probability, 85–94
  frequency concept of, 86–89
  and justification. *See* Justification, and probability

logical conception of, 89–90, 91–93
subjective conception of, 90–93
and truth, 91–92, 93
Proper functioning, 6–7, 184–185
Putnam, Hilary, 148(n1), 207, 231(n9), 240*b*

Quine, W. V. O., 9, 22(nn12, 13), 60, 69(n10), 109, 118, 119, 122(nn11, 19), 144, 149(n3), 178, 203, 203(n7), 240*b*
Quinton, Anthony, 240*b*

Radford, Colin, 22(n14), 34–36, 38, 43, 44(nn12–14), 240*b*
Rationalist epistemology, 49
Reason
life of, 41
and metamind, 40
and preference, 40
See also Reasonableness
Reasonableness, 126–128
as a condition of knowledge, 15
and expected utility, 145–148
as a primitive, 144
and probability, 144–148
and trustworthiness, 142–44
and truth, 147–148
See also Evaluation system; Lottery paradox
Reasoning system, 127
defined, 130
Reductionism. *See* Phenomenalism
Reichenbach, Hans, 96(n13), 240*b*
Reid, Thomas, 71, 73, 95(n1), 96(n3), 143, 148(n2), 206, 230, 231(n5), 240*b*
Reliabilism, 184, 190–193, 223
Reliability
and belief, 74–75
and independent information, 74–75
and justification game, 198–200
and trustworthiness, 192–195
See also Externalism
Rescher, Nicholas, 148, 230, 231(n4), 240*b*
Rosenberg, Jay F., 121, 144, 149(n3), 240*b*
Ross, Glen, 230, 241*b*
Roth, Michael, 174, 230, 240*b*, 241*b*
Russell, Bertrand, 21, 22(n9), 23(n21), 68, 101, 121(n4), 241*b*
Ryle, Gilbert, 22(n8), 241*b*

Savage, W. C., 241*b*
Schick, Frederick, 122(n16)
Schlick, Moritz, 241*b*
Self-justification, 45, 71, 84
contingent, 83–85
and foundationalism, 48
as incorrigibility. *See* Incorrigibility
and independent information, 72–73, 75–76, 84
and necessary truth, 79–83
and semantics, 76. *See also* Semantic foundationalism
See also Basic belief, Fallible foundations; Incorrigibility; Semantic foundationalism
Sellars, Wilfrid F., 96(n8), 101, 103, 109, 118, 119, 121, 121(n3), 122(nn11, 18), 144, 149(n3), 241*b*
Semantic field, 78
Semantic foundationalism, 77–83
Shoemaker, Sydney, 241*b*
Shope, Robert K., 174, 241*b*
Siegel, Harvey, 241*b*
Simplicity, 118–119
Skepticism, 2, 18, 128–129, 205–209
and conceptual change, 206–207
and contextualism, 218–219
and deductive closure, 214–217, 220–222
and externalism, 180–181, 214–217
and fallible knowledge, 208–209
and ignorance. *See* Agnoiology
and meaning, 77–79
and the possibility of error, 206–208
and relevant alternatives, 181, 215, 217
and trustworthiness, 209–212, 220

Slaght, Ralph, 241*b*
Sosa, Ernest, 69(n7), 95, 174(n1), 205, 209, 213, 231(nn1, 11, 13), 232(nn14, 17), 241*b*
Stine, G. C., 232(n21), 241*b*
Stroud, Barry, 230, 231(n4), 242*b*
Subjective justification. *See* Personal justification
Swain, Marshall, 21, 174, 196–197, 203, 204(n22), 230, 240*b*, 242*b*
Swartz, Robert, 235*b*

Tarski, Alfred, 27–28, 29–30, 43, 43(n1), (44n2), 242*b*
Thalberg, Irving, 174(n4), 242*b*
Transparency, 61–62. *See also* Opacity.
Truncellito, David, 204(n21)
Trustworthiness, 140
  of acceptance. *See* Principle of trustworthiness
  of preference, 141–142
  and reasonableness, 139
  of reasoning, 140–141
  and reliability, 192–195, 211
  and truth, 224, 227–228
  and virtue, 209–212
  See also Principle of trustworthiness
Truth, 26–31
  absolute theory of, 12. *See also* Truth, general theory of
  complete theory of. *See* Truth, general theory of.
  as a condition of knowledge, 11–12
  and correspondence, 31
  disquotational theory of, 29, 30
  general definition of. *See* Truth, general theory of
  general theory of, 26–31
  minimal theory of. *See* Truth, disquotational theory of
  and paradox, 27–31
  See also Justification; Principle of trustworthiness
Tweedale, Jonathan, 175(n11), 242*b*

Ullian, Joseph, 240*b*
Ultra justification, 153–154. *See also* Undefeated justification.
Ultra justification game
  and the Gettier problem, 154–158, 160–162
  and the isolation objection, 163–165
  and the truth connection, 163–165
Ultrasystem, 154
  defined, 168
Undefeated justification, 152–173
  defined, 169, 171
Unger, Peter, 230, 231(n4), 242*b*

Van Cleve, James, 84, 95, 96(n11), 242*b*
Vander Nat, Arnold, 21, 239*b*
Virtue epistemology, 209. *See also* Trustworthiness.

Wagner, Carl, 239*b*
Warfield, Ted, 230, 236*b*
Will, Frederick L., 242*b*
Williams, Michael, 205, 222, 231(n2), 232(n22), 242*b*
Wittgenstein, Ludwig, 242*b*

Zagzebski, Linda, 205, 209, 231(nnl, 11), 242b
Ziff, Paul, 121, 121(n5), 242*b*